

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Studi sul Patrimonio Culturale – Cultural Heritage Studies

Ciclo XXXII

Settore Concorsuale: 11/A4 - Scienze del libro e del documento

Settore Scientifico Disciplinare: M-STO/08 - Archivistica, bibliografia e biblioteconomia (primario);
INF/01 - Informatica (secondario)

Impiego del Web Semantico per lo sviluppo e la consultazione di
archivi musicali.

Un caso di studio sulla storia e la documentazione del
melodramma italiano: l'archivio *Corago*.

Presentata da: **Paolo Bonora**

Coordinatore Dottorato
Prof. Raffaele Savigni

Supervisore
Prof. Angelo Pompilio

Esame finale anno 2020

Abstract

Lo studio presentato in questa tesi ha affrontato tre ambiti: la sperimentazione delle tecnologie semantiche per produrre una rappresentazione della storia e dei documenti relativi al melodramma; l'individuazione di metodologie di analisi dei contenuti della base di conoscenza per estrarre nuova informazione; l'esplorazione di nuove modalità di accesso e consultazione dei contenuti. La sperimentazione ha impiegato un insieme di tecnologie informatiche con particolare attenzione al Web Semantico, ai Linked Data e al Natural Language Processing.

Il caso di studio è costituito dall'archivio del melodramma italiano *Corago* che raccoglie un vasto patrimonio di informazioni relative al repertorio, alla cronologia degli spettacoli e ai libretti. I contenuti dell'archivio sono stati trasposti in una base di conoscenza mediante l'adozione di due ontologie formali di riferimento: CIDOC Conceptual Reference Model e FRBR Object Oriented.

La sperimentazione ha prodotto i seguenti risultati: (i) la realizzazione di un modello concettuale in grado di esprimere efficacemente il patrimonio informativo dell'archivio; (ii) la definizione di una metodologia di rappresentazione dell'informazione del dominio orientata alla ricezione dei contenuti; (iii) l'elaborazione di un modello descrittivo dei personaggi; (iv) la definizione di una metodologia per l'individuazione di relazioni tra i testi dei libretti; (v) la valutazione funzionale del modello concettuale attraverso una applicazione di consultazione dell'archivio.

Nel complesso, lo studio evidenzia come gli strumenti oggetto della sperimentazione offrano significative potenzialità nella realizzazione di sistemi per la gestione dell'informazione storica e documentale che viene sempre più prodotta e distribuita in forma digitale. L'evoluzione verso sistemi in grado di gestire e scambiare informazioni sulla base della semantica dei rispettivi modelli si offre come soluzione efficace alla gestione di contenuti multiformi e complessi come quelli relativi alla storia del melodramma.

Parole chiave: rappresentazione della conoscenza, estrazione dell'informazione, *linked data*, storia del melodramma, CIDOC CRM, FRBR, archivi digitali.

“Vedi caro, in fondo tu sei un *informanista*:
né informatico, né umanista!”

Vincenzo Di Massimo

Indice

1	Introduzione	1
1.1	Premessa	1
1.2	Tema della ricerca	3
1.3	Metodo	3
1.4	Contesto della sperimentazione.....	5
1.5	Requisiti generali per il sistema sperimentale	6
1.5.1	Requisiti di base	6
1.5.2	Ricerche dei contenuti testuali.....	7
1.5.3	Ricerche spazio-tempo	7
1.5.4	Analisi delle relazioni tra i testi.....	8
1.5.5	Requisiti non funzionali	8
1.6	Struttura dei contenuti della tesi	9
1.6.1	Capitolo 1: Introduzione	9
1.6.2	Capitolo 2: Rappresentazione della conoscenza di dominio	9
1.6.3	Capitolo 3: Analisi dei contenuti.....	9
1.6.4	Capitolo 4: Estensione del patrimonio informativo	9
1.6.5	Capitolo 5: Strategie di accesso ai contenuti	9
1.6.6	Capitolo 6: Conclusioni	9
2	Rappresentazione della conoscenza di dominio	10
2.1	Stato dell'arte nella rappresentazione della conoscenza mediante ontologie formali	10
2.2	Dal modello relazionale al modello concettuale a semantica esplicita.....	13
2.2.2	Dal database relazionale al <i>triple store</i>	15
2.2.3	Il <i>mapping</i>	22
2.2.4	Analisi della qualità del <i>dataset</i>	24
2.2.5	Risultati	25
2.3	Specializzazione del modello concettuale: il Corago Semantic Model.....	27
2.3.1	Rappresentazione delle responsabilità autoriali	30
2.3.2	Rappresentazione del ruolo dell'interprete	32
2.4	L'organizzazione dei contenuti nella prospettiva funzionale FRBR.....	33
2.5	Navigazione guidata dalla semantica di dominio	35
2.5.1	Introduzione	35
2.5.2	Entità ed attributi: dalla morfologia del grafo alla rappresentazione funzionale	35
2.5.3	Modello delle proprietà di navigazione.....	37
2.5.4	Costruzione delle viste utente	39
2.6	Indicizzazione dei contenuti testuali	45
2.6.1	Articolazione del requisito funzionale.....	45
2.6.2	Analisi tecnica ed implementazione dell'indice	46
3	Analisi dei contenuti	56
3.1	Lo spazio ed il tempo	57
3.1.1	La dimensione geografica del melodramma	57
3.1.2	La dimensione temporale	58
3.1.3	Analisi coordinate spazio-tempo	59
3.2	I personaggi come dimensione di analisi del testo drammatico	60

3.2.1	I contenuti dell'archivio sui personaggi.....	62
3.2.2	Il modello descrittivo del personaggio d'opera.....	63
3.2.3	Analisi della tradizione dei testi attraverso i personaggi	68
3.2.4	Conclusioni preliminari ed evoluzioni future	76
4	Estensione del patrimonio informativo.....	77
4.1	Stato dell'arte delle metodologie di analisi automatica dei contenuti	77
4.1.1	Analisi NLP e tagging morfosintattico.....	78
4.1.2	Estrazione dell'informazione non supervisionata basata su <i>pattern</i>	78
4.2	Estrazione dell'informazione dai contenuti esistenti	79
4.2.1	I casi di studio: descrizioni di personaggi ed ambientazioni.....	79
4.2.2	Analisi NLP dei contenuti testuali.....	80
4.2.3	Annotazione morfosintattica: UD in RDF	82
4.2.4	Estrazione basata su regole	83
4.2.5	Estrazione basata su fonti esterne	90
4.3	Allineamento ed inclusione di contenuti da fonti LOD.....	93
4.3.1	Georeferenziazione dei toponimi.....	93
4.3.2	Associazione Opere al RISM	95
5	Strategie di accesso ai contenuti.....	97
5.1	Ricerche testuali	97
5.2	Ricerche per dimensioni spazio-tempo e di dominio	102
5.2.1	Ricerche spazio-tempo	102
5.2.2	Ricerche basate sui personaggi	104
5.3	Ricerca avanzata per attributi	106
5.4	Rappresentazione e navigazione multidimensionali.....	109
5.4.1	Gli elenchi di risultati.....	111
5.4.2	Le schede delle entità.....	112
5.4.3	Rappresentazione su cartografia.....	117
5.5	Valutazione del sistema.....	120
5.5.1	Struttura del questionario di valutazione.....	121
5.5.2	Modalità di somministrazione e rilevamento dei risultati	123
5.5.3	Analisi quantitativa dei risultati.....	124
5.5.4	Valutazione delle risposte aperte.....	132
5.5.5	Considerazioni complessive sui risultati ottenuti.....	133
6	Conclusioni	134
6.1	Rappresentazione dei contenuti	134
6.2	Analisi dei contenuti.....	134
6.3	Accesso ai contenuti.....	135
	Prefissi utilizzati	137
	Allegati digitali	137
	Appendici.....	138
A.	Mappe concettuali per la trasposizione dei contenuti.....	138
A.1.	<i>Mapping</i> dei Responsabili	138
A.1.1.	<i>Mapping</i> degli attributi.....	139
A.1.2.	<i>Mapping</i> delle entità associate	139
A.2.	<i>Mapping</i> dei Brani	140

A.2.1.	<i>Mapping</i> delle entità associate	142
A.3.	<i>Mapping</i> degli Spettacoli.....	146
A.3.1.	<i>Mapping</i> degli attributi.....	147
A.3.2.	<i>Mapping</i> delle entità associate	147
A.4.	<i>Mapping</i> dei Personaggi.....	150
A.4.1.	<i>Mapping</i> degli attributi.....	151
A.4.2.	<i>Mapping</i> delle entità associate	151
A.5.	<i>Mapping</i> dei Titoli.....	153
A.5.1.	<i>Mapping</i> degli attributi.....	153
A.5.2.	<i>Mapping</i> delle entità associate	154
A.6.	<i>Mapping</i> dei Legami	155
A.6.1.	<i>Mapping</i> degli attributi.....	156
A.6.2.	<i>Mapping</i> delle entità associate	156
A.7.	<i>Mapping</i> delle Sezioni	157
A.7.1.	<i>Mapping</i> degli attributi.....	158
A.7.2.	<i>Mapping</i> delle entità associate	158
A.8.	<i>Mapping</i> dei Documenti.....	159
A.8.1.	<i>Mapping</i> degli attributi.....	160
A.8.2.	<i>Mapping</i> delle entità associate	160
A.9.	<i>Mapping</i> degli esemplari.....	166
A.9.1.	<i>Mapping</i> degli attributi.....	167
A.9.2.	<i>Mapping</i> delle entità associate	167
A.10.	<i>Mapping</i> della Bibliografia.....	169
A.10.1.	<i>Mapping</i> degli attributi.....	170
A.10.2.	<i>Mapping</i> delle entità associate	170
A.11.	Modello relazionale Corago.....	173
B.	Architettura del sistema sperimentale.....	174
B.1.	Architettura del sistema	174
B.2.	Applicazione di consultazione	175
B.3.	Indicizzazione full-text.....	175
B.4.	Annotazione ed analisi dei contenuti testuali.....	176
C.	Componenti software.....	177
C.1.	Indicizzazione full-text.....	177
C.1.1.	Costruzione dell'indice: il metodo 'createFullTextIndex(String)'	177
C.2.	Strumenti di analisi NLP.....	182
C.2.1.	Generatore training set	182
C.2.2.	Annotatore NLP-RDF.....	184
C.3.	Utilizzo delle proprietà di navigazione	188
	Bibliografia.....	195

1 Introduzione

L'obiettivo del lavoro che viene qui presentato è verificare l'efficacia delle tecnologie informatiche per la rappresentazione formale della conoscenza applicate alla gestione dei contenuti relativi alla storia del melodramma.

La sperimentazione è stata condotta adottando la prospettiva funzionale dell'utente finale e in questa ottica ha voluto indagare le potenzialità di questi strumenti attraverso l'allestimento di un sistema sperimentale di consultazione dei contenuti dell'archivio del melodramma italiano *Corago* (Pompilio et al., 2005).

I risultati, sul piano metodologico, tecnico e funzionale, che vengono presentati nei capitoli seguenti vogliono essere un contributo d'ambito disciplinare alla riflessione circa l'efficacia dell'adozione di queste tecnologie per la raccolta, analisi e diffusione del vasto patrimonio informativo d'interesse storico.

1.1 Premessa

La storiografia dello spettacolo deve necessariamente confrontarsi con la natura effimera del proprio oggetto di studio. Una volta spente le luci della ribalta, la vera natura di quanto andato in scena è persa per sempre. Restano solo dei frammenti documentali, testimoni parziali di una unità composita¹. Nel caso dello spettacolo operistico, la sua natura di "oggetto complesso" nel quale interagiscono e si compenetrano testi di natura letteraria, musicale, coreografica e visiva, fa sì che la varietà documentale sia massima. Se le partiture ed i libretti testimoniano i contenuti della creazione artistica di compositore e librettista; bozzetti, figurini, disposizioni sceniche e appunti di regia documentano la realizzazione dell'opera.

Alla varietà dei materiali documentali corrisponde la varietà delle istituzioni preposte alla loro conservazione: biblioteche, archivi e musei. I diversi mandati istituzionali e l'intrinseca specificità delle tipologie di documenti conservati hanno prodotto nel tempo una moltitudine di modelli descrittivi e norme catalografiche diverse. Il patrimonio informativo prodotto da queste istituzioni è confluito in sistemi informativi autonomi e con criteri di accesso eterogenei. Un utente interessato a conoscere gli eventi performativi o consultare la documentazione del teatro d'opera deve essere consapevole della frammentazione delle informazioni che sta ricercando, deve saper utilizzare i diversi sistemi informativi disponibili e sobbarcarsi l'onere di ricostruire il nesso tra i diversi contenuti rintracciati.

Per mettere l'utente nella condizione di accedere in modo diretto al patrimonio informativo dello spettacolo, superando la disseminazione dei documenti in sistemi informativi diversi, è necessario realizzare un modello concettuale capace di gestire in modo coordinato descrizioni dei contenuti e dei documenti utilizzati come fonti ed infine di esplicitarne le relazioni e rappresentarne la trama. Data questa complessità, è quindi necessario adottare un modello concettuale in grado di gestire in modo coordinato l'informazione storica e quella documentale nelle sue diverse tipologie.

I progetti *RADAMES*² prima, e *Corago* poi, hanno cercato di dare una risposta a questa esigenza. La sfida è quella di costruire un sistema in grado di trattare in modo omogeneo «l'informazione storica, bibliografica e documentale relativa al teatro d'opera, un patrimonio culturale circoscritto ma complesso, caratterizzato da una documentazione quanto mai ricca e multiforme» (Bianconi et al., 2004, p. 354).

¹ Si rinvia a (Bianconi et al., 2004) per una trattazione esaustiva del tema di cui qui si riporta una sintesi.

² Per una introduzione all'implementazione del modello concettuale promossa dal progetto si rinvia a (Pompilio et al., 2005).

Il modello concettuale elaborato in *RADAMES* utilizza come riferimento quanto previsto dai *Functional Requirements for Bibliographic Records* (FRBR) promosso dall'International Federation of Library Associations and Institutions (IFLA)³. FRBR introduce un cambio di paradigma nel modo di rappresentare l'informazione bibliografica. La novità rilevante del modello concettuale introdotto da FRBR, rispetto alla tradizione delle pratiche catalografiche tradizionali, sta nell'aver spostato il punto di attenzione del processo di catalogazione dai criteri di compilazione della scheda catalografica alla identificazione della natura del contenuto informativo presente in una scheda catalografica, distinguendo chiaramente la descrizione del contenuto intellettuale e la descrizione dell'oggetto materiale.

La natura composita del record bibliografico viene esplicitata in FRBR in quattro livelli descrittivi distinti e organizzati gerarchicamente. A partire dal livello descrittivo relativo all'oggetto materiale, il singolo *item*, si identifica un secondo livello relativo al suo prototipo editoriale (*manifestation*), segue il terzo livello nel quale vengono descritti i contenuti di una specifica edizione di un'opera (*expression*) per culminare nel vertice rappresentato dall'opera (*work*), intesa come ideazione concettuale capostipite comune di tutte le forme attraverso viene descritta ai livelli inferiori.

In base all'articolazione sui quattro livelli previsti dal modello FRBR, il modello concettuale predisposto per il teatro d'opera dal progetto *RADAMES*, ed adottato per la realizzazione dell'archivio *Corago*, organizza le informazioni intorno a tre nuclei principali: il repertorio delle opere, la cronologia degli eventi e l'archivio dei documenti. Ciascuno dei tre contenitori impiega un modello descrittivo disegnato sulle caratteristiche peculiari degli oggetti contenuti. L'archivio documentale censisce l'informazione bibliografica relativa agli oggetti materiali: libretti, documenti, edizioni e manoscritti. La Cronologia riporta le informazioni storiche relative agli spettacoli nei quali sono state rappresentate le opere liriche descritte nel Repertorio.

Un modello concettuale così strutturato ha consentito la realizzazione di un sistema informativo capace di raccogliere in un ambiente omogeneo l'informazione storica e documentale dello spettacolo operistico. In questo modo offre un punto di accesso unico ai diversi materiali ricomponendone la frammentazione sul tavolo da lavoro virtuale dello studioso.

Il modello di per sé non riduce la complessità dell'informazione sottostante, ma offre la possibilità di definire strategie efficaci per la fruizione di questo patrimonio anche in relazione a profili di utenza diversificati. Nella valutazione dell'efficacia del sistema informativo è determinante la flessibilità nell'elaborazione degli strumenti per l'accesso e la rappresentazione dell'informazione.

Nel corso degli ultimi vent'anni sono emerse e si sono affermate un insieme di tecnologie informatiche che hanno introdotto la possibilità di definire la semantica espressa dai modelli concettuali in termini formali. Questo consente di affiancare ai dati una definizione formale del modello descrittivo del dominio di riferimento espressa attraverso le cosiddette *logiche descrittive* (*description logic* o DL). In questa tipologia di sistemi i due piani convivono all'interno della stessa base informativa e condividono la stessa famiglia di linguaggi formali attraverso cui vengono rappresentati modelli e contenuti.

Tra le ontologie formali sviluppate nell'ambito dell'informazione bibliografica e museale rilevanti sono il Conceptual Reference Model promosso dal CIDOC⁴ (CIDOC CRM) e FRBR Object-Oriented (FRBROO) promosso dall'IFLA. In particolare, la seconda rappresenta l'armonizzazione del modello concettuale dedicato ai beni museali espresso nel CIDOC CRM ed il modello FRBR come riferimento per l'ambito dell'informazione bibliografica. L'obiettivo è la realizzazione di un modello unitario che permetta l'interscambio di informazioni

³ Si rinvia a (IFLA Study Group, 1998).

⁴ Comitato Internazionale per la Documentazione (CIDOC) dell'International Council of Museums (ICOM).

tra i due ambiti e la possibilità di sviluppare sistemi informativi interoperabili a disposizione degli utenti interessati ad accedere a contenuti condivisi (Bekiari et al., 2015).

L'introduzione delle tecnologie semantiche, oltre ad aprire la prospettiva di un'interoperabilità a livello concettuale tra sistemi informativi eterogenei, offre la possibilità di operare all'interno di un sistema descrittivo coerente ed omogeneo. Diventa quindi possibile sperimentare diverse modalità di accesso e rappresentazione dell'informazione contenuta all'interno del sistema con l'obiettivo di rispondere all'evolversi delle richieste degli utenti e per estenderne il patrimonio informativo.

La verifica dell'efficacia di questi strumenti è particolarmente rilevante in un archivio come *Corago* il cui patrimonio informativo, essendo in grado di associare informazioni sul repertorio, la cronologia e bibliografia, costituisce il punto di incontro tra gli interessi degli studiosi del melodramma e i bibliotecari.

La complessa articolazione dei contenuti, la ricchezza dei riferimenti verso fonti informative e collezioni digitali esterne e i diversi profili degli utenti, rendono l'archivio *Corago* un caso di studio particolarmente fertile.

1.2 Tema della ricerca

Lo studio qui presentato si è posto come obiettivo primario la sperimentazione dell'adozione delle tecnologie semantiche al fine di verificarne l'efficacia nel gestire, analizzare ed arricchire i contenuti attinenti alla storia del melodramma.

In particolare, si è voluto indagare:

- l'efficacia delle ontologie formali nel rappresentare la conoscenza storica relativa al melodramma;
- l'applicazione di strumenti per l'analisi dei contenuti e l'estrazione di nuova conoscenza;
- le ricadute sul piano funzionale nella gestione dell'informazione attraverso l'impiego di questi strumenti;
- la fruibilità dell'informazione così rappresentata.

1.3 Metodo

Per rispondere agli obiettivi della ricerca è stato adottato un approccio sperimentale basato sulla elaborazione e verifica delle metodologie, degli strumenti e delle tecnologie sottoposti ad indagine. Per la valutazione della efficacia funzionale delle soluzioni individuate si è adottato un approccio comparativo andando a confrontare i risultati ottenuti nel singolo contesto sperimentale rispetto al termine di paragone costituito dall'attuale sistema *Corago* o, dove necessario, dallo stato dell'arte. Gli esiti dei diversi impianti sperimentali sono stati sottoposti a verifiche di natura quantitativa e qualitativa a seconda dei casi (Hevner et al., 2004).

Per ciascun tema di indagine si è quindi proceduto con:

- la definizione degli obiettivi generali del sistema sperimentale;
- la definizione degli obiettivi specifici derivati dai requisiti generali;
- la definizione del contesto applicativo e ricognizione degli strumenti utili alla sperimentazione;
- l'implementazione della fase sperimentale;
- la verifica dei risultati.

Per la verifica dei risultati conseguiti sono state impiegate due distinte metodologie:

- questionari utente per la valutazione dell'implementazione dei requisiti⁵;
- l'efficacia degli strumenti di analisi è stata valutata in relazione all'avanzamento della conoscenza che sono stati in grado di produrre⁶.

⁵ Lo strumento adottato e le metodologie applicate vengono illustrati nel paragrafo 5.5.

⁶ Si rinvia ai paragrafi 3.2.3, 4.2.4 e 4.2.5 per l'illustrazione dei casi in cui questa metodologia di valutazione è stata applicata.

1.4 Contesto della sperimentazione

La ricerca si colloca in un percorso avviato negli ultimi quindici anni durante i quali lo sviluppo delle tecnologie per la gestione di sistemi informativi per la catalogazione e descrizione della conoscenza nell'ambito dei Beni Culturali ha prodotto strumenti in grado di affiancare ai dati catalografici una loro rappresentazione formale attraverso modelli concettuali. L'introduzione di questi strumenti ha esteso la possibilità di disegnare sistemi informativi dedicati alla rappresentazione della conoscenza con la possibilità di apportare due novità significative.

La prima è rappresentata dalla flessibilità nella definizione dei modelli descrittivi, nella relativa semplicità di revisione e nell'adeguamento degli stessi all'arricchirsi dei contenuti. Infatti, maggiori sono i livelli di astrazione in cui si è in grado di articolare il modello, più agevole sarà operare su specifiche parti per raffinarlo; allo stesso tempo, quanto maggiore risulterà la sua analiticità, tanto maggiore sarà la granularità con cui si può intervenire sul dato e proporre piani di lettura differenziati in funzione del profilo utente.

La seconda novità deriva dalla possibilità di formalizzare, attraverso le ontologie formali, la semantica dei modelli descrittivi fornendo così una definizione puntuale dei concetti del dominio attraverso uno strumento intellegibile sia all'attore umano che alle macchine. Il primo vantaggio che ne scaturisce è l'introduzione di una reale interoperabilità sul piano della semantica dei contenuti tra sistemi informativi operanti su modelli concettuali diversi. Questa possibilità rende attuabile l'estensione dei rispettivi perimetri informativi superando i limiti fisici del singolo catalogo attraverso l'introduzione di relazioni con risorse provenienti da altri contesti.

Inoltre, operando su modelli concettuali strutturati gerarchicamente, questi sistemi sono in grado di rappresentare l'informazione offrendo, in fase di consultazione, molteplici livelli di astrazione concettuale costruiti a partire dal livello descrittivo di base dei dati. Sfruttando questa articolazione, gli esperti di dominio potranno adeguare la rappresentazione delle informazioni in funzione dei requisiti funzionali delle diverse tipologie di utenti.

Ne deriva che le metafore visuali di navigazione del patrimonio informativo del catalogo possono beneficiare sia della molteplicità dei livelli di rappresentazione del dato che dell'introduzione di dimensioni di analisi derivanti dall'estensione dello stesso grazie all'interoperabilità con fonti esterne. In questo senso si rendono sperimentabili metafore in grado di rendere compiutamente l'articolazione dei contenuti e delle relazioni tra di essi.

Allo stesso tempo, l'interoperabilità verso altri sistemi permette l'introduzione di nuove dimensioni di analisi non contemplate nel modello originario. In questo modo informazioni provenienti da fonti diverse, potenzialmente eterogenee per contenuti e finalità, possono essere integrate e sfruttate dagli utenti come chiavi di accesso ai contenuti del catalogo. L'esplicitazione della semantica del modello descrittivo diventa quindi il prerequisito per l'integrazione dell'archivio in questo ecosistema.

All'interno di questo contesto la sperimentazione è stata avviata a partire dalla definizione dei requisiti funzionali che un archivio musicale sviluppato su tecnologie semantiche dovrebbe soddisfare.

1.5 Requisiti generali per il sistema sperimentale

1.5.1 Requisiti di base

L'identificazione dei requisiti funzionali per il sistema sperimentale è stata svolta a partire dall'analisi delle caratteristiche funzionali del sistema 'tradizionale' di consultazione dell'archivio *Corago*⁷. L'obiettivo originario del sistema 'tradizionale' è fornire uno strumento di supporto alla ricerca sulla storia materiale dello spettacolo operistico e allo studio della tradizione del testo letterario del melodramma italiano sulla base della raccolta di informazioni ricavate da fonti primarie e secondarie (Bonora e Pompilio, 2019).

Per la consultazione dell'archivio vengono proposte all'utente tre funzionalità di ricerca distinte: una per il repertorio, una per la cronologia degli spettacoli ed una terza per il catalogo dei documenti. Ciascuna offre un sottoinsieme di attributi specifici per ciascuna delle tre entità utilizzabili come criteri per interrogare il sistema. I risultati della ricerca vengono presentati in forma tabellare attraverso una selezione di attributi che descrivono le caratteristiche. A partire dalla singola riga di risultato si accede alla scheda che riporta una descrizione dettagliata relativa all'entità selezionata comprensiva dei riferimenti alle risorse correlate. Ad esempio, all'interno della scheda di un'opera viene presentato l'elenco delle relative rappresentazioni e dei libretti associati.

Il sistema prevede quindi che il punto di ingresso sia necessariamente legato alla scelta preventiva di una delle tre modalità di ricerca e che l'esplorazione delle relazioni esistenti tra le diverse entità avvenga solo a valle della selezione di uno specifico risultato. Sia gli attributi utilizzati per descrivere i risultati negli elenchi, sia le schede descrittive di ciascuna delle tre tipologie di entità, sono quindi necessariamente vincolati alla tipologia di entità essendo un sottoinsieme selezionato degli attributi descrittivi previsti per la singola tipologia di entità.

All'utente viene proposta una modalità di accesso verticale basata sulle caratteristiche specifiche di ciascuna delle tre entità fondamentali: opere, eventi e libretti. La modalità di esplorazione delle relazioni orizzontali esistenti tra esse è così vincolata alla scelta di una di esse come tipologia di partenza. Questa impostazione sottolinea la centralità di opere, eventi e libretti come punti privilegiati di accesso ai contenuti dell'archivio.

In termini pratici questo comporta che l'utente debba formulare la sua strategia di ricerca scegliendo da quale tipologia di entità partire in funzione della applicabilità dei criteri che intende utilizzare agli attributi che il sistema prevede per ciascuna tipologia di entità. L'ammissibilità o meno di un attributo come criterio di ricerca lo qualifica implicitamente come chiave discriminante per le entità censite nei tre indici.

In questa prospettiva, diventa rilevante per il sistema di ricerca individuare a priori ed in forma esplicita quali siano i criteri in base a cui ciascuna tipologia di entità deve poter essere ricercata [REQ1]⁸.

Allo stesso tempo, nella consultazione degli elenchi dei risultati, gli attributi utilizzati per rappresentarli sono gli unici elementi che permettono all'utente di valutare la congruità e la rilevanza del risultato ottenuto in funzione dei criteri impostati. Questi attributi sono necessari per identificare il singolo risultato all'interno dell'insieme proposto dal sistema. Questa loro funzione li distingue da tutti gli altri attributi presenti nel modello descrittivo di ciascuna entità in quanto ne definiscono l'identità.

Ne consegue che gli attributi in grado di definire l'identità delle istanze di una specifica classe devono essere esplicitamente contrassegnati come attributi identitari [REQ2].

⁷ Il sistema è consultabile all'indirizzo web: <http://corago.unibo.it> (verificato il 27/09/2019).

⁸ Nel seguito della trattazione e nei capitoli successivi si farà riferimento ai requisiti utilizzando la sigla [REQ n] dove n rinvia al numero progressivo del requisito in questo paragrafo.

La struttura della scheda che viene presentata all'utente in fase di navigazione è a sua volta una rappresentazione del modello descrittivo legata alla tipologia di entità rappresentata. Gli attributi che la compongono e il modo in cui vengono organizzati e presentati ne veicolano il contenuto informativo e influenzano la sua ricezione. La disposizione dei contenuti nella scheda è quindi una modalità di rappresentazione dell'informazione specifica del dominio.

Ne deriva la necessità di definire per ciascun attributo del modello descrittivo la rilevanza e la conseguente modalità di presentazione all'interno della struttura della scheda [REQ3].

1.5.2 Ricerche dei contenuti testuali

Le diverse tipologie di dato vengono organizzate durante la catalogazione secondo la struttura prevista dal modello descrittivo. Informazioni testuali, numeriche e date confluiscono nei metadati specifici di ciascuna entità del dominio. È così che la stringa "Venezia" può indicare la località di prima rappresentazione di un'opera, il luogo di edizione di un libretto, la città natale di un autore o l'ambientazione di un quadro scenico. Utilizzato come criterio di ricerca, una porzione di testo può essere simultaneamente applicata a più attributi descrittivi a prescindere dalla tipologia di entità ricercata e dal relativo significato dell'attributo considerato.

L'utente può così effettuare una ricerca *full-text* sull'intero archivio senza dover specificare il tipo di entità ed il relativo attributo. Si tratta quindi di una ricerca testuale libera, svincolata dalla struttura del modello descrittivo delle singole entità. Per rispondere a questo requisito i contenuti testuali di ciascuna risorsa dovranno confluire in un indice testuale che consenta di individuare le corrispondenze col testo ricercato. Le corrispondenze dovranno consentire all'utente di valutare la congruità e la rilevanza del risultato ottenuto rispetto agli obiettivi della ricerca.

Un sistema in grado di rispondere a questo requisito deve permettere di specificare a priori quali contenuti testuali debbano essere indicizzati, impiegare l'indice per la ricerca in modalità *full-text* e produrre come risultato un elenco di entità con l'indicazione di quali attributi contengano il testo ricercato [REQ4].

1.5.3 Ricerche spazio-tempo

Così come ipotizzato per i contenuti testuali, è possibile immaginare ricerche che abbiano per oggetto dimensioni comuni a più tipologie di entità. In un dominio centrato sulle arti performative le dimensioni spazio e tempo hanno una particolare rilevanza. La carriera di un interprete, la fortuna di un'opera, la produzione di un teatro diventano più leggibili se rappresentate lungo la dimensione spazio-temporale. Le stesse tre entità principali del dominio hanno una loro stretta collocazione rispetto ad entrambe le dimensioni⁹.

La possibilità di associare i contenuti dell'archivio ad entrambe le dimensioni consente di effettuare ricerche combinate spazio-tempo. La collocazione deve però posizionare le tipologie di entità rispetto a una definizione univoca e condivisa delle due dimensioni. Ad esempio, per poter collocare uno spettacolo lungo una linea temporale, si dovrà considerare la data di rappresentazione mentre, per il corrispondente libretto, quella di edizione. Le due entità verranno quindi collocate sulla stessa dimensione temporale ma con due precisioni diverse: il giorno, mese o stagione per l'evento, l'anno per il libretto.

⁹ Per eventi e libretti la collocazione è fisiologica, per l'opera è più complesso. In prospettiva FRBR la datazione di un'opera intesa come *Work* è legata alla sua data di concezione da parte dell'autore. Tuttavia, va precisato che per le opere liriche convenzionalmente si utilizza come riferimento temporale la data di prima rappresentazione. Similmente può essere derivata la collocazione spaziale, sarebbe più significativo derivarla dall'ambientazione dell'opera stessa.

La collocazione rispetto allo spazio dovrà necessariamente derivare dalla relazione che un'entità ha con un luogo, normalmente rappresentato dal suo toponimo. La collocazione rispetto allo spazio sarà quindi rappresentata dalle coordinate geografiche associabili a quest'ultimo.

Se, a partire dal patrimonio informativo esistente, risulterà possibile attestare i contenuti rispetto alle dimensioni spazio-tempo, sarà possibile per l'utente utilizzarle sia come criteri di ricerca che come modalità di rappresentazione dei risultati. Infatti, a fronte di una interrogazione combinata di periodo e area geografica, il sistema potrà rappresentarne i risultati su cartografia.

Per raggiungere questi risultati, saranno analizzati i contenuti dell'archivio per derivarne la collocazione rispetto alle due coordinate [REQ5]. Verrà individuata la modalità di rappresentazione delle due dimensioni, le modalità del loro impiego come criteri di ricerca e come debbano essere presentati all'utente i risultati ottenuti in base ad esse [REQ6].

1.5.4 Analisi delle relazioni tra i testi

Spazio e tempo sono dimensioni generali che assumono particolare rilievo rispetto a contenuti di natura storica. Se è vero che il loro utilizzo come dimensioni di analisi introduce nuove modalità di accesso al patrimonio informativo dell'archivio, si può ipotizzare che l'individuazione di dimensioni specifiche del dominio possa offrire nuovi strumenti per indagare le fonti ed i relativi contenuti.

Una prima ipotesi di lavoro riguarda i personaggi. Di solito i libretti riportano, un elenco dei nomi dei personaggi e ciascuno di essi è corredato da una breve didascalia che ne descrive sinteticamente attributi e carattere. Sebbene le didascalie possano variare per estensione e contenuto costituiscono comunque un elemento utile per identificare il personaggio e collocarlo nel contesto dell'azione drammatica. In questa prospettiva, i personaggi risultano essere uno degli elementi qualificanti i contenuti del testo letterario dell'opera.

Per verificare quest'ipotesi, è necessario indagare l'efficacia del personaggio come dimensione di analisi in grado di qualificare i contenuti di un'opera e quindi di individuare eventuali relazioni che possono intercorrere tra i diversi testi letterari. Nell'archivio *Corago* il personaggio è trattato come entità autonoma distinta dall'opera in cui compare. Per ciascun personaggio viene riportato: denominazione, descrizione e la relazione con l'opera.

Per utilizzare il personaggio come chiave di accesso al repertorio è necessario disegnare un modello in cui far confluire gli elementi descrittivi identificati. In base ad essi si potrà tentare di confrontare i personaggi sulla base delle caratteristiche a prescindere dal nome. Eventuali affinità o identità tra personaggi di opere diverse possono risultare indicatori utili per individuare relazioni tra libretti.

Il sistema informativo dovrà essere in grado di rappresentare il personaggio come entità caratterizzata da una serie di attributi e offrire strumenti per la navigazione del repertorio a partire da essi [REQ7].

1.5.5 Requisiti non funzionali

Per soddisfare i requisiti espressi finora è necessario organizzare ed integrare l'informazione presente nell'archivio sia attraverso l'analisi dei contenuti già disponibili, sia attraverso l'introduzione di nuove informazioni derivate da fonti esterne.

Questo richiede la completa trasposizione del patrimonio informativo dell'archivio in un modello concettuale che permetta di:

1. rappresentare i contenuti con un grado di analiticità coerente con la semantica del dato in fonte;

2. estendere il perimetro informativo per includere nuove definizioni di entità sia interne che esterne al dominio;
3. offrire la possibilità di organizzare in modo gerarchico i livelli di conoscenza.

Si dovrà quindi verificare che l'insieme di questi requisiti venga soddisfatto dalla trasposizione del modello concettuale originale dell'archivio in un modello a semantica esplicita basato su ontologie formali [REQ8].

1.6 Struttura dei contenuti della tesi

La struttura della tesi che descrive il percorso della sperimentazione effettuata in base ai requisiti generali del sistema sperimentale è la seguente:

1.6.1 Capitolo 1: Introduzione

Viene presentato il tema della ricerca e le premesse metodologiche, il contesto e lo stato dell'arte ed indicati i requisiti generali per la sperimentazione.

1.6.2 Capitolo 2: Rappresentazione della conoscenza di dominio

Si illustrano i metodi e gli strumenti per la rappresentazione dei contenuti dell'archivio attraverso modelli a semantica esplicita. Si affrontano gli aspetti funzionali di tale rappresentazione e gli strumenti individuati per renderla fruibile.

1.6.3 Capitolo 3: Analisi dei contenuti

Si presentano le dimensioni di analisi sperimentate in questo studio. Vengono presentati i relativi modelli descrittivi adottati. Vengono illustrate le tecniche impiegate per l'organizzazione dei contenuti in funzione di esse ed i risultati ottenuti in termini di efficacia nell'accesso al patrimonio informativo dell'archivio.

1.6.4 Capitolo 4: Estensione del patrimonio informativo

Vengono descritti i procedimenti per l'inclusione di contenuti provenienti da fonti informative esterne all'archivio. Vengono illustrate le metodologie e gli strumenti per l'estrazione dell'informazione dai contenuti esistenti. Viene presentata la valutazione dei risultati ottenuti.

1.6.5 Capitolo 5: Strategie di accesso ai contenuti

Si illustrano le soluzioni individuate per l'accesso ai contenuti. Viene illustrata la applicazione web per la consultazione dell'archivio. Viene presentata la valutazione della copertura funzionale raggiunta rispetto ai requisiti generali.

1.6.6 Capitolo 6: Conclusioni

Vengono sintetizzate le conclusioni della ricerca svolta ed illustrate alcune ulteriori direzioni in cui possa essere condotta.

2 Rappresentazione della conoscenza di dominio

Il capitolo illustra il processo di trasposizione del patrimonio informativo dell'archivio *Corago* dalla forma relazionale utilizzata dall'applicazione di catalogazione al formato Resource Description Framework (RDF)¹⁰ che costituirà la base di conoscenza oggetto della sperimentazione funzionale. Questa nuova versione dell'archivio andrà sotto la denominazione di *Corago Linked Open Data* (Corago LOD), ovvero sotto forma di *dataset* costruito in base ai principi Linked Data.

Il processo è stato realizzato in tre fasi: trasposizione dei contenuti in formato RDF; rappresentazione dei contenuti espressi in forma di grafo; indicizzazione ai fini della ricerca e consultazione.

La prima fase ha consentito di verificare compiutamente la fattibilità della trasposizione dei contenuti dal modello concettuale originale di tipo relazionale ad un modello a semantica esplicita basato su due ontologie formali¹¹ di riferimento: il *Conceptual Reference Model* di CIDOC e la versione Object Oriented del modello FRBR. Il *dataset* ottenuto è stato pubblicato come risorsa *open data*¹².

Nella seconda fase sono state sviluppate rappresentazioni ritenute efficaci dal punto di vista funzionale dei contenuti espressi nella nuova forma. La complessa morfologia che la struttura dei dati ha assunto ha sollecitato l'elaborazione di un approccio alla rappresentazione in grado di coniugare l'efficacia funzionale alla coerenza formale del modello semantico di riferimento.

In fine, nella terza fase sono state definite le modalità di indicizzazione dei contenuti testuali in grado di rispondere ai requisiti di accesso anche in considerazione della possibilità di utilizzare risorse provenienti da fonti informative esterne. Il metodo di indicizzazione elaborato è in grado di restituire oltre alla corrispondenza con le chiavi di ricerca anche il loro contesto di provenienza.

Il risultato del processo di trasposizione del patrimonio informativo dell'archivio ha prodotto la base di conoscenza su cui sono state svolte le attività di analisi dei contenuti descritte nei capitoli 3 e 4 e l'allestimento dell'applicazione di consultazione descritta nel capitolo 5. Nel complesso l'allestimento della base di conoscenza ha consentito di rispondere a quanto previsto dal [REQ8].

2.1 Stato dell'arte nella rappresentazione della conoscenza mediante ontologie formali

A partire dagli anni Duemila, il modello del Web Semantico (Semantic Web - SW) ha sollecitato la progettazione di basi dati la cui architettura coniuga il dato grezzo al relativo modello concettuale espresso attraverso ontologie formali. Nell'ambito della catalogazione in sistemi informativi bibliografici e museali, si prospetta un possibile scenario in cui i dati raccolti nei processi di catalogazione confluiscono in vere e proprie basi di conoscenza che costituiscono la base per realizzare un ecosistema fondato sull'interoperabilità semantica: ovvero i Linked Data (Bizer et al., 2011).

L'esplicitazione della semantica delle entità di un dominio, delle relazioni che intercorrono tra di esse ed il mondo esterno permette di superare il tradizionale limite delle applicazioni di consultazione basate su

¹⁰ Per una introduzione al formato vedi (McBride, 2004).

¹¹ L'analisi di fattibilità della trasposizione, l'individuazione degli strumenti e la prototipazione della soluzione è stata oggetto della tesi di laurea: "Trasposizione di un indice repertoriale del melodramma nel paradigma Linked Open Data" di Paolo Bonora, relatore prof. Paolo Nesi, discussa nel A.A. 2014/2015 presso l'Università di Firenze. Parte dei contenuti dei paragrafi 2.1, 2.2, 2.3 riportano gli esiti di quello studio.

¹² Il *dataset* in formato RDF N-TRIPLE è pubblicato attraverso la piattaforma Zenodo ed accessibile attraverso il DOI: 10.5281/zenodo.3377586

modelli descrittivi a 'semantica implicita'¹³. Ad esempio, se si osservano i cataloghi bibliografici delle biblioteche che adottano lo stesso standard descrittivo ISBD, questi richiedono il possesso di competenze specifiche in merito alla concreta implementazione per poter essere realmente sfruttati a pieno. All'estremo opposto si collocano i motori di ricerca generalisti di contenuti web che, a partire da semplici chiavi testuali inserite dall'utente, propongono risultati o percorsi di navigazione basati su euristiche implicite e, di norma, proprietarie. L'impiego di ontologie descrittive per la definizione formale della semantica del modello di dominio e la rappresentazione di entità e relazioni in RDF consentono la realizzazione di sistemi che si collocano in posizione intermedia tra i due estremi. Da un lato si ha accesso al dato con il massimo controllo sulla granularità della semantica del contenuto ricercato, dall'altra si possono costruire percorsi di navigazione guidata dei contenuti dell'archivio che prescindano da una conoscenza a priori di come questi siano organizzati. Questi percorsi possono essere definiti sfruttando la stratificazione di diversi livelli di astrazione concettuale costruiti a partire dal livello descrittivo di base fino a quello di massima granularità. L'esperto di dominio potrà sfruttare questa articolazione su più livelli per adeguare le modalità di fruizione ai requisiti funzionali delle diverse tipologie di utenti, siano essi attori umani o altri sistemi informativi.

Al tempo stesso, l'interoperabilità tra sistemi (che in prospettiva Linked Data richiede l'allineamento semantico tra i relativi modelli concettuali) permette di estendere il perimetro di ciascun indice oltre i confini del proprio modello e apre la possibilità di includere dimensioni originariamente non presenti nella base di conoscenza. Possono essere così integrate informazioni provenienti da fonti diverse, potenzialmente eterogenee per contenuti e finalità, in grado di fungere da nuove chiavi di accesso ai contenuti già presenti nell'archivio.

L'esplicitazione della semantica del modello concettuale dell'archivio non solo costituisce il prerequisito per l'integrazione dell'indice come LOD, ma permette anche di modellare specifici paradigmi di interrogazione e rappresentazione del dato operando sui diversi livelli di astrazione definiti a partire dalle ontologie formali utilizzate per rappresentare il dominio.

In ambito museale, l'International Council of Museums (ICOM) nel 1994 elabora il CIDOC Relational Data Model (RDM), un modello generale di rappresentazione di beni museali orientato all'intercambio informativo tra istituzioni. Nella seconda metà degli anni Novanta viene affrontata la sua trasformazione in modello *object oriented* che si concluderà con l'introduzione del CIDOC Conceptual Reference Model (CIDOC CRM) divenuto uno standard dell'International Organization for Standardization (ISO) nel 2006¹⁴. Già durante il processo di approvazione del CRM nei primi anni Duemila, è avvertita l'esigenza di trovare una forma di raccordo tra i modelli descrittivi utilizzati dal mondo museale e quello delle biblioteche. Si tratta di armonizzare i due modelli in modo che le semantiche dei concetti condivisi risultino allineate (Doerr M. e LeBoeuf P., 2007). Come per il CIDOC CRM, anche per FRBR si deve partire dalla sua forma relazionale per derivarne una ontologia formale che prende il nome di FRBR *Object Oriented* (FRBRoo) la cui prima versione viene pubblicata nel 2008.

Oggi queste due ontologie hanno raggiunto un discreto grado di diffusione e la comunità di riferimento comincia a ragionare compiutamente sugli aspetti tecnici e procedurali derivanti dal loro impiego come

¹³ Con la locuzione 'semantica implicita' si fa riferimento alla mancanza di una descrizione formale del significato delle entità e dei relativi contenuti rappresentati dal modello.

¹⁴ Il percorso di standardizzazione, cominciato nel 2000, si conclude nel settembre del 2006 quando il modello CRM viene adottato come standard numero ISO 21127:2006 successivamente aggiornato nel 2014 (ISO 21127:2014).

ontologie di riferimento per *dataset* LOD¹⁵; per questa ragione sono naturali ontologie di riferimento per la trasposizione dei contenuti dell'archivio *Corago* sotto forma di *dataset* RDF.

Alla fine degli anni Duemila, hanno cominciato ad essere resi disponibili *dataset* LOD con contenuti di ambito musicale. Sono state introdotte ontologie specifiche per il dominio come The Music Ontology¹⁶ e sono stati sperimentati allineamenti con *dataset* generalisti come DBPedia¹⁷ (Raimond Y. e Sandler M., 2008). Queste iniziative sono state oggetto di interesse sia da parte dei produttori di contenuti che di gestori di cataloghi e repertori alla ricerca di strumenti per la diffusione del patrimonio informativo censito nei propri sistemi. Un esempio è il progetto promosso dalla British Broadcasting Corporation (BBC) settore Radio nel 2009 per la definizione di un modello di aggregazione della propria programmazione radiofonica che sfrutta l'allineamento del proprio catalogo con *dataset* esterni (Kobilarov G. et al., 2009). L'obiettivo dichiarato è quello di utilizzare le informazioni ricavate per offrire ulteriori chiavi di accesso ai propri materiali. Un secondo esempio può essere rappresentato dal progetto MIMESIS (Bartolini I. et al., 2012) che ha tentato di definire un approccio basato sugli strumenti del Web Semantico per l'utilizzo del Repertorio della Poesia Italiana in Musica 1500-1700 (RePIM)¹⁸. L'obiettivo in questo caso era proporre all'utente la ricostruzione della tradizione dei testi indicizzati lungo l'asse orizzontale (analisi dell'articolazione del testo) e temporale (versioni e relazioni dei testi) attraverso una rappresentazione strutturata in base all'ontologia FRBROO (*ibidem*, p. 50).

In entrambi i casi l'introduzione di una rappresentazione formalizzata della semantica dei contenuti viene impiegata con più finalità. La prima è la convergenza verso un modello descrittivo condiviso con l'esterno delle proprie risorse che diventano comparabili con quelle provenienti da altre fonti definite in base allo stesso modello. Il risultato ricercato è la costruzione di uno strato descrittivo omogeneo e coerente sul quale costruire logiche di accesso, aggregazione o associazione delle entità. La possibilità di mettere in relazione oggetti esterni con risorse locali permette di estendere il perimetro informativo di partenza e di includere riferimenti e nuove dimensioni di analisi. Come corollario ne deriva una estensione dell'interoperabilità tra le diverse basi di conoscenza che risultano così legate oltre che da modelli comuni anche da istanze di relazioni tra risorse su cui costruire percorsi di navigazione condivisi percepiti come unitari dall'utente finale. Un ultimo beneficio derivante dalla trasposizione di modelli a semantica implicita verso ontologie e formati dati standard è la garanzia di manutenibilità e conservazione del patrimonio informativo censito. Oltre al fisiologico tasso di obsolescenza tecnologica che può mettere a rischio l'accessibilità futura delle basi di conoscenza che impieghino strumenti informatici, deve essere considerato anche il modello concettuale alla base come fattore sottoposto a manutenzione e documentazione. La trasformazione dei dati da un formato proprietario ad un formato aperto come quello RDF li rende accessibili e gestibili con diversi strumenti riducendo il rischio di obsolescenza del formato originale e dei software in grado di utilizzarlo. Allo stesso tempo, l'esplicitazione del modello concettuale attraverso ontologie come CIDOC CRM e FRBROO ne

¹⁵ Ad esempio, il tema relativo alla corretta formazione degli Uniform Resource Identifier per le risorse censite nei *dataset* con risorse di ambito museale (cfr.: http://www.cidoc-crm.org/URIs_and_Linked_Open_Data.html).

¹⁶ Sito web: <http://musicontology.com/> (consultato il 15/09/2019).

¹⁷ Sito web: <http://wiki.dbpedia.org/> (consultato il 15/09/2019).

¹⁸ Il progetto di allineamento tra Early Music Online (EMO) e l'Electronic Corpus of Lute Music (ECOLM) può rappresentare un ulteriore esempio di applicazione degli strumenti del Web Semantico al repertorio musicale (Crawford et al., 2014).

garantisce l'intelligibilità dei contenuti da attori diversi delegando ad esse anche l'onere di mantenerne viva la competenza necessaria per l'utilizzo¹⁹.

2.2 Dal modello relazionale al modello concettuale a semantica esplicita

L'obiettivo di questa fase è ottenere la trasposizione dei contenuti del modello relazionale *Corago* in formato RDF esplicitandone la semantica attraverso l'adozione delle due ontologie di riferimento: FRBR Object Oriented e CIDOC CRM²⁰ (Doerr, M., 2009). In questo passaggio è stato definito l'allineamento tra la semantica delle classi previste dalle due ontologie e la semantica delle entità presenti nel modello relazionale dell'archivio. Per individuare la corrispondenza tra entità originali e classi di destinazione è stato necessario ricostruire il valore semantico dei contenuti tenendo conto non solo della natura dei dati ma anche della loro origine: ovvero del processo di catalogazione che li ha prodotti.

Il modello descrittivo impiegato dall'archivio *Corago* rappresenta infatti l'ultima evoluzione di una pluridecennale attività di catalogazione e di estensione del modello originale *RADAMES* elaborato nell'ambito delle arti performative ed applicato, in particolare, al melodramma (A. Pompilio et al., 2005). Sebbene il disegno generale del modello adottasse fin dall'origine l'impianto proposto dalle linee guida FRBR, questo è venuto via via specializzandosi per rispondere alle esigenze specifiche della storiografia musicale. Proprio l'alto livello di specializzazione del dominio ha richiesto l'introduzione di alcune estensioni ai modelli definiti dalle due ontologie di riferimento. Queste sono state raccolte nell'ontologia denominata *Corago Semantic Model*²¹ (Corago SM). Il modello concettuale della base di conoscenza utilizza pertanto le due ontologie di riferimento CIDOC CRM e FRBRoo e il Corago SM come ontologia specialistica di dominio.

In ultimo ultima considerazione di natura architettuale. Nel processo di trasposizione è stato realizzato un sistema che consentisse di minimizzare l'impatto sul DBMS di origine. Si è proceduto quindi a selezionare degli strumenti per il *mapping* e l'estrazione che garantissero un sufficiente grado di disaccoppiamento rispetto alla sorgente dati e fossero già disponibili e di libero utilizzo.

2.2.1.1 Il processo di trasposizione

Il processo di trasposizione può essere descritto come la composizione di una trasformazione fisica e strutturale del dato e la contestuale introduzione di un ulteriore livello descrittivo che ne va a specificare morfologia e semantica. La trasformazione opera sul piano fisico una riorganizzazione dei dati dalla tupla del formato relazionale alle triple del formato RDF. Questa porta nella forma di tripla RDF, strutturata come terna ordinata di soggetto, predicato e oggetto, i metadati descrittivi delle entità *Corago* originariamente organizzati in tabelle. Le tabelle strutturano il dato sotto forma di ennupla formata da coppie etichetta-valore dove l'etichetta identifica e descrive il contenuto del campo. Da questa si deve desumere, in relazione col modello descrittivo entità-relazioni di cui la tabella fa parte, il significato del contenuto del campo.

Nella trasformazione della coppia in triple RDF il contenuto del campo diviene il soggetto di una serie di triple che, mediante predicati diversi, devono essere in grado di rappresentare l'informazione desumibile dal

¹⁹ Il tema della conservazione del patrimonio informativo non verrà qui espressamente trattato. Resta però un elemento ineludibile per quanto complesso: si compone infatti di problematiche che vanno dagli aspetti manutentivi delle architetture e dei formati delle basi di dati per garantirne l'accessibilità, fino all'esplicitazione e conservazione dei criteri catalografici che hanno prodotto la base di conoscenza in assenza dei quali, la mera fruibilità tecnica del dato risulta poco significativa.

²⁰ Per entrambe le ontologie verrà utilizzata la cosiddetta versione "Erlangen" mantenuta dall'Università di Erlangen-Nuremberg (sito web: <http://erlangen-crm.org/about>).

²¹ La definizione del Corago SM in linguaggio OWL è pubblicata attraverso la piattaforma GitHub all'indirizzo: <https://github.com/paolobonora/Corago-SM>.

modello originario. I predicati utilizzati, la cui semantica viene definita dalle ontologie di riferimento, possono qualificare il dato da diversi punti di vista: dalle caratteristiche fisiche o formali (si tratta di numeri, stringhe o date specificate secondo un particolare formato) fino alla natura dell'entità rappresentata. Il predicato previsto per questo tipo di 'asserzione' è la proprietà *rdf:type* che identifica la classe di appartenenza di ciascun dato.

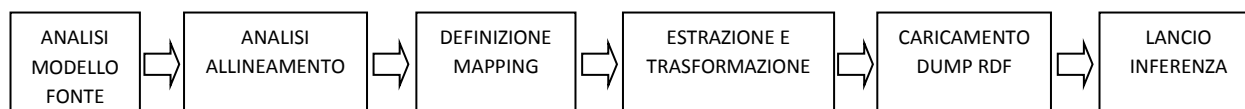
L'estrazione delle entità Corago dal contesto fisico della base dati di origine deve necessariamente utilizzare un meccanismo di identificazione univoca delle singole istanze. Nel formato RDF, l'entità sarà rappresentata nelle triple attraverso un identificativo univoco: l'Uniform Resource Identifier (URI). La singola entità potrà trovarsi in posizione di soggetto o di oggetto. La singola tripla potrà mettere in relazione due entità, oppure specificare un valore di un attributo attraverso un nodo di tipo letterale (*literal*).

Ciascuna entità capostipite, ad esempio un'opera, sarà quindi rappresentata da un grafo di triple. La morfologia del grafo risultante dalla trasposizione dei dati e la sua capacità di rappresentare efficacemente i contenuti estratti dalla fonte originale dipendono dall'espressività delle ontologie formali utilizzate come riferimento (Allenmang e Hendler, 2011, p. 22).

La trasposizione dei contenuti non può ovviamente prescindere dalla conoscenza della semantica del modello di origine per poter individuare il corretto allineamento con le ontologie utilizzate. Nel caso di *Corago* la scelta di FRBRoo come ontologia di riferimento risulta naturale data la comune origine FRBR. Nel corso dell'analisi per definire l'allineamento si comunque proceduto alla puntuale verifica della corrispondenza tra contenuti informativi estratti e definizioni destinate a rappresentarli.

Concretamente la trasformazione avviene sulla base di una associazione tra il campo della tabella di origine e la tripla di destinazione, ovvero l'insieme di triple che rappresenteranno le proprietà del valore nel grafo RDF in base al modello espresso dall'ontologia di riferimento. Questa attività di '*mapping*' è materialmente rappresentata da una definizione formale ed eseguibile del processo di estrazione e serializzazione del dato dal DBMS fonte al file RDF. L'esecuzione dell'estrazione e trasformazione del dato produce il cosiddetto '*dump RDF*'. Questo livello è il livello al quale i contenuti della base di conoscenza sono asseriti direttamente (*stated*) utilizzando i predicati previsti dalle ontologie di riferimento utilizzate dal *mapping*. In questa forma il *dataset* può essere caricato in un database a grafo (o *triple store*) in grado di gestire le procedure di caricamento, estrazione e ricerca, ma anche di inferenza. L'inferenza farà sì che, a partire dai tipi asseriti per ciascuna risorsa in funzione del *mapping*, questa venga associata a tutti quei tipi che rappresentano una generalizzazione derivabile in base alle definizioni presenti nelle ontologie di cui viene fatta l'inferenza.

Il processo di trasformazione viene quindi ad essere articolato in più fasi: analisi del modello concettuale di origine; analisi dell'allineamento rispetto alle ontologie di riferimento; definizione del *mapping*; esecuzione dell'estrazione; caricamento nel *repository* e lancio dell'inferenza. Lo schema 1 rappresenta graficamente le macro-fasi di processo che devono essere implementate.

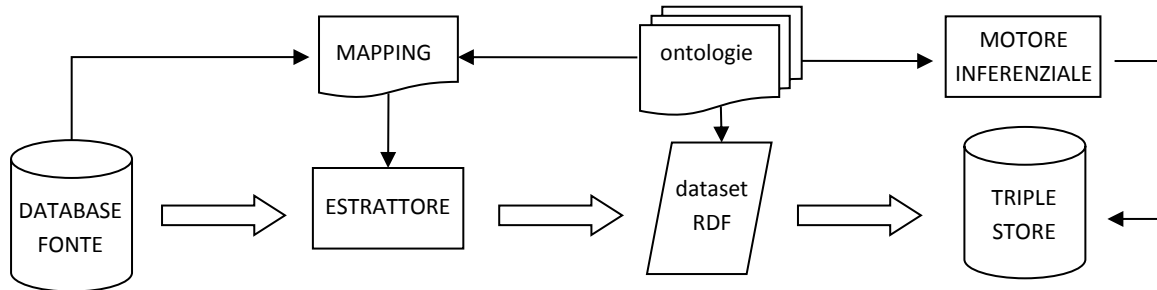


Schema 1: Fasi del processo di trasposizione

Dal punto di vista dell'architettura software necessaria per eseguire il processo di trasformazione si può adottare una suddivisione funzionale di massima articolata in: componenti per la definizione del *mapping* e

l'esecuzione dell'estrazione del *dump*, un *triple store* da utilizzare come *repository* e motore inferenziale per il lancio dell'inferenza sul *dataset* RDF, strumenti accessori ed infrastruttura per la connettività delle diverse componenti.

Lo schema 2 rappresenta questa prima suddivisione ed il relativo flusso dati.



Schema 2: Architettura di *mapping*, estrazione e caricamento del *dataset*

2.2.2 Dal database relazionale al *triple store*

Per operare la trasposizione dal tracciato relazionale al formato RDF si è proceduto in due fasi: la prima relativa all'esplicitazione della semantica del modello concettuale Corago, la seconda relativa al *mapping* fisico della fonte con il modello definito come target. Definita la metodologia di allineamento tra i due modelli si è proceduto alla scelta degli strumenti, all'implementazione e messa a punto del processo di estrazione dei dati.

2.2.2.1 Metodologia adottata

Per la definizione e successiva implementazione del processo di conversione ed estrazione, si è proceduto in quattro fasi distinte:

1. allineamento del modello relazionale Corago alle ontologie di riferimento;
2. design e l'implementazione del *mapping* fisico;
3. processo di estrazione e caricamento;
4. verifica della coerenza e qualità del dato esposto.

Il processo di estrazione è stato realizzato in maniera iterativa. In particolare, nei primi due passaggi quando sono stati riscontrati vincoli implementativi all'allineamento o la necessità di specializzare il modello concettuale attraverso l'ontologia Corago SM.

2.2.2.1.1 L'allineamento del modello relazionale Corago alle ontologie di riferimento

Il modello concettuale Corago ed il processo di catalogazione che ha portato alla costruzione dell'attuale base di conoscenza è frutto di un percorso lungo ed articolato. Nel tempo il modello descrittivo, nelle diverse implementazioni che si sono succedute, ha permesso di catalogare diverse forme tipologie di documenti caratteristici delle arti performative: dai libretti d'opera (Cipollone B., 2009), alle registrazioni audio, fino ai costumi teatrali (Bonora P., 2010). Pensato fin dall'origine come implementazione del modello FRBR, compresi elementi derivati da FRAD²², pur mantenendosi coerente nell'impianto complessivo, ne rappresenta una significativa specializzazione. Questa specializzazione risponde alla specificità dei processi catalografici propri della ricerca storiografica di settore. Ad esempio, la capacità descrittiva della relazione che intercorre tra autore e opera (la cosiddetta 'responsabilità autoriale'), l'individuazione di legami tra opere, la descrizione della loro articolazione interna (segmentazione) hanno richiesto estensioni specifiche nel modello descrittivo.

In passato erano già stati tentati approfondimenti in merito alla possibilità di tradurre alcune entità catalogate col modello *RADAMES* nel modello CIDOC CRM (Bonora et al., 2006) senza però giungere ad una concreta implementazione del processo di trasposizione²³. Date queste premesse, il primo passo è stato definire il perimetro del modello Corago da utilizzare come fonte per la futura base di conoscenza con la finalità di esporre un insieme coerente ed omogeneo di entità che andranno a costituire il *dataset* RDF, evitando di includere entità peculiari o numericamente non significative²⁴.

2.2.2.1.2 Definizione del perimetro del *mapping*

Nell'identificare le entità, i loro attributi e relative relazioni sono stati considerati due fattori: la rappresentatività e rilevanza rispetto al dominio di partenza e la centralità rispetto al modello semantico di destinazione. Il primo permette di esporre i tratti salienti della conoscenza di dominio raccolta in *Corago*, il secondo consente una corretta e coerente fruibilità da sistemi e soggetti terzi rispetto ad un modello non specializzato.

La tabella 2.1 riporta la denominazione della tabella fonte di *Corago*²⁵, una descrizione sintetica dell'entità principale che vi viene descritta e le classi o proprietà di CRM o FRBRoo derivate da essa.

²² FRAD: Functional Requirements for Authority Data è il modello concettuale di riferimento proposto dall'IFLA per descrizione delle entità censite negli authority file. Nasce come complemento a FRBR per la definizione di soggetti, thesauri, tassonomie e relazioni a supporto delle entità in esso definite.

²³ Sperimentazioni che restano tra le prime esperienze nell'impiego di CRM nell'ambito delle arti performative (Doerr M., 2009).

²⁴ Il modello Corago comprende diverse aree tematiche o funzionali, quali quelle relative alla gestione degli oggetti digitali, che in questa sede si è deciso di non includere nel perimetro delle entità interessate dall'attività di trasposizione in RDF anche in considerazione del livello di catalogazione cui erano giunte.

²⁵ Vengono riportate solo le entità e le relative tabelle principali, sono state omesse alcune tabelle di relazione o di servizio comunque necessariamente incluse nel *mapping*.

	Entità	Tabella Corago	Contenuto	Classi derivate
1	Responsabile	RESPONS	Anagrafica di tutte le responsabilità censite nel catalogo	E39_Actor
2	Brano	BRANI	Catalogo dei brani, dalle opere complete alle singole parti	F1_Work e F2_Expression nelle relative sottoclassi
3	Titolo	TITOLI	Indice generali dei titoli	E35_Title
4	TitoloBrano	BRA_TIT	Associazioni di altri titoli ai Brani	P102_has_title
5	ResponsabileBrani	BRA_AUT	Attestazione delle responsabilità dei Brani	C2_Actor_Role
6	Spettacolo	SPETTACOLI	Catalogo degli eventi (Spettacoli)	F31_Performance
7	Personaggio	PERSONAG	Catalogo dei Personaggi	F38_Character
8	BraniSpettacoli	SPET_BRA	Articolazione contenuto degli Spettacoli	F25_Performance_Plan
9	Responsabilità BraniSpettacoli	SPET_BRA_RB	Articolazione delle responsabilità autoriali dei Brani negli Spettacoli	C2_Actor_Role
10	Responsabilità Interpreti	SPET_BRA_CAN	Individuazione dei ruoli interpreti e Personaggi per gli Spettacoli	C2_Actor_Role
11	Responsabilità Spettacolo	SPET_RES1	Individuazione altre responsabilità per Spettacoli	C2_Actor_Role
12	Legami	LEGAMI	Indice delle relazioni tra Brani	C5_Link
13	Documenti	DOCUMENTI	Catalogo delle fonti documentali	F3_Manifestation_Product_Type
14	Bibliografia	BIBLIOGRAFIA	Indice della bibliografia di riferimento	E31_Document
15	RISM	RISM	Indice delle Biblioteche	E42_Identifier
16	Collocazioni	SWT_DOC_RISM	Attestazione dei Documenti presenti in Biblioteche	F5_Item
17	Sezioni	SEZ1	Suddivisioni in atti, quadri scenici o parti delle opere	F1_Work

Tabella 2.1

Una rappresentazione, in forma semplificata, della parte del modello Corago interessata dal *mapping* viene illustrata dal diagramma 2.1.



1. le relazioni tra le diverse entità sono spesso qualificate da attributi (raccolti in classi di associazione) e, in alcuni casi, sono ternarie: ovvero mettono in relazione più entità;
2. le fonti documentarie vengono esplicitate dalla relazione tra le entità di dominio (Brani, Responsabili e Spettacoli) e i documenti utilizzati come fonte.

Il modello FRBR, nella sua versione relazionale (FRBR-ER), può essere invece sintetizzato nel diagramma 2.2.

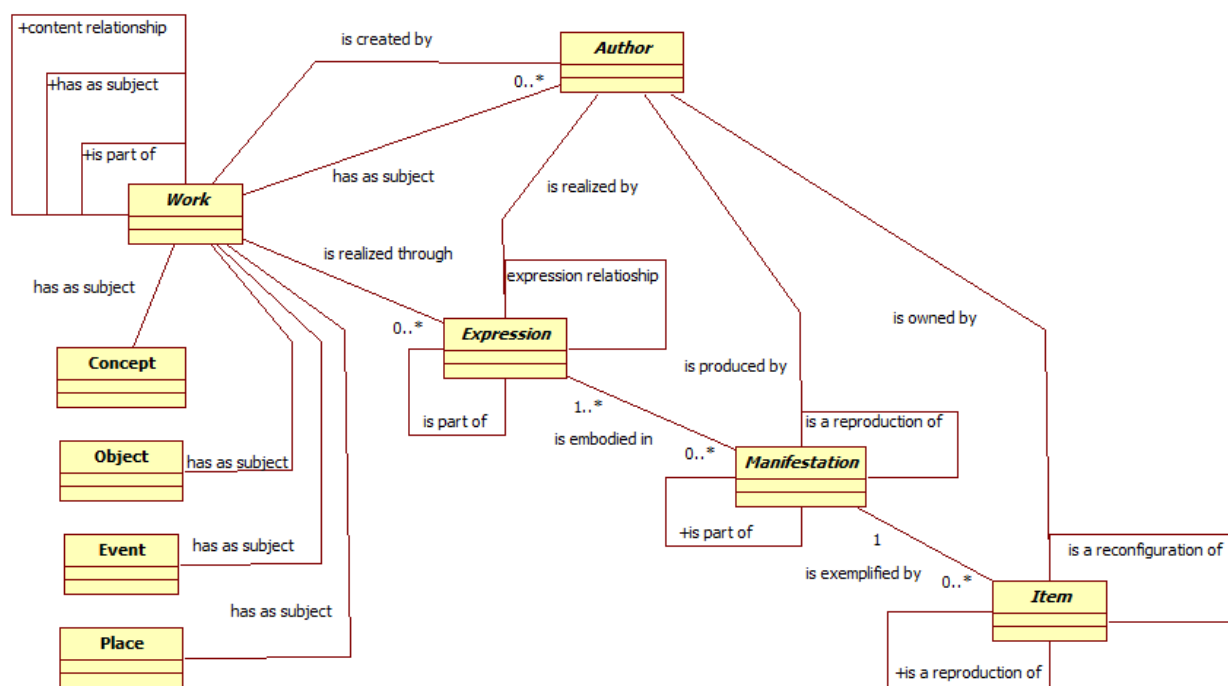


Diagramma 2.2: Il modello logico FRBR-ER

Per quanto sia evidente la derivazione del primo dal secondo, sussistono alcune differenze di una qualche rilevanza. Il modello FRBR-ER nasce per descrivere il processo che porta dall'ideazione dell'opera fino alla sua manifestazione fisica, il tutto in una prospettiva fortemente orientata a descrivere il processo editoriale dell'opera letteraria o artistica. In particolare, viene operata una distinzione tra Opera (*Work*) e sua Espressione (*Expression*), identificando nella prima il contenuto astratto dell'opera stessa, nella seconda la sua realizzazione formalizzabile (indipendentemente dalla forma che andrà ad assumere).

Per definire le equivalenze tra entità Corago ed entità FRBRoo si è proceduto suddividendo il modello originale in macro-aree aventi come centro l'entità principale e la sua tabella fonte. Per ciascuna entità/tabella se ne sono analizzati i contenuti ed il processo di catalogazione che la ha popolata in modo da ricostruire i diversi possibili significati ed individuare di conseguenza le classi adeguate ad esprimerli. Nella selezione delle classi di destinazione si è tenuto conto, non solo della semantica espressa dalle loro definizioni e del relativo allineamento a quella della fonte, ma anche della posizione che vengono ad occupare nel complesso sistema descrittivo derivante dalla composizione delle due ontologie. Ad esempio, in alcuni casi si è preferito utilizzare la classe definita nel CIDOC CRM, pur esistendo una specializzazione in FRBRoo, prediligendo la generalizzazione del concetto (dove non sussistessero vincoli che rendessero la seconda espressamente non utilizzabile). Dove le due ontologie non offrivano gli strumenti necessari a rappresentare efficacemente il contenuto si è dovuto procedere ad una loro estensione nell'ontologia Corago SM.

In particolare, si è intervenuti introducendo:

1. una serie di specializzazioni di E55_Type per rappresentare alcuni tipi Corago particolarmente significativi e che si ritiene utile poter individuare come tali in usi futuri;

2. l'estensione del modello della responsabilità autoriale rappresentato in CRM/FRBROO dalla coppia di proprietà `crm:P14_carried_out` e la proprietà di proprietà `crm:P14.1_in_the_role_of`²⁶.

La trattazione delle scelte fatte nell'allineamento dei modelli verrà approfondita per ciascuna delle macro-aree assieme alle specificità implementative nei paragrafi relativi al *mapping* riportati nell'Appendice A – Mappe concettuali.

2.2.2.2 Criticità e soluzioni nell'allineamento dei modelli concettuali

Tra le diverse considerazioni relative all'allineamento concettuale tra il modello relazionale Corago e le ontologie adottate come riferimento del *dataset* Corago LOD è opportuno evidenziarne due per la rilevanza e gli esiti che assumono sul piano metodologico.

La prima è relativa al passaggio tra il piano delle opere a quello dell'espressione rappresentata in *Corago* dagli eventi. L'ontologia FRBROO utilizza la classe `frbroo:F31_Performance` per rappresentare lo spettacolo. Questa classe è una specializzazione della classe CIDOC CRM `crm:E7_Activity`, nel modello FRBROO quindi le istanze della classe rappresentano eventi storici propriamente detti.

La relazione con il piano delle espressioni è mediata dalla classe `frbroo:F25_Performance_Plan` che, in quanto sottoclasse della `frbroo:F22_Self-Contained_Expression`, rappresenta l'insieme delle direttive necessarie alla messa in scena. Nel caso di performance teatrali e musicali queste comprendono i testi letterari e musicali oggetto della specifica rappresentazione (Bekiari et al., 2015, p. 57). Il nesso tra opera ed evento è quindi stato rappresentato attraverso il grafo: `frbroo:F1_Work` – `frbroo:R9_is_realized_in` – `frbroo:F25_PerformancePlan` – `frbroo:R25_performed` – `frbroo:F31_Performance`. Nell'implementazione di questa struttura si è scelto però di creare una istanza di `frbroo:F25_PerformancePlan` per ogni attestazione di contenuto riportata nella tabella associativa SPET_BRA del modello relazionale Corago.

La scelta è motivata dalla necessità di reificare il singolo record fonte a livello di espressione di un contenuto autonomo. Ciascuno di essi fa riferimento ad un'opera specifica, ha un titolo proprio ed un insieme di responsabilità caratteristiche. Questa soluzione, se da un lato forza l'unitarietà del concetto di Performance Plan che rimane pur sempre ricavabile dall'unione delle istanze che collegano lo spettacolo alle opere che lo compongono, dall'altra rappresenta efficacemente il livello descrittivo dell'informazione presente in *Corago*.

Il secondo aspetto è legato all'espressività delle due ontologie nel rappresentare l'articolazione della relazione tra autore e opera già citata che, per essere superata, ha richiesto l'estensione del modello concettuale con l'introduzione di una nuova classe e di relative proprietà associative.

Questi due casi rappresentano il confine tra possibilità di 'interpretare' il modello concettuale delle ontologie di riferimento e la necessità di estendere la loro capacità di rappresentare i fatti del dominio d'interesse. Nel fare questo, resta valida l'indicazione di procedere per specializzazione offrendo la possibilità di accedere ai contenuti attraverso la rappresentazione generalizzata.

²⁶ Il meccanismo individuato nei due modelli ha due limiti significativi: non è implementabile in RDF, almeno senza ricorrere a soluzioni non deprecate, non consente l'estensione di relazioni ternarie (ad es. come quella tra cantante, ruolo e personaggio interpretato). Per una analisi approfondita del problema e delle relative ipotesi di soluzione si veda (Alexiev, 2012).

2.2.2.3 Strumenti

Stante la necessità di allestire un *repository* dei contenuti dell'archivio in formato RDF e il relativo *endpoint* SPARQL²⁷ si è proceduto nella selezione degli strumenti che ne permettessero l'allestimento tenendo fermi alcuni requisiti di base:

1. essere di libero utilizzo;
2. non richiedere modifiche al file MDB²⁸ di origine;
3. coprire i requisiti di espressività necessaria ad implementare il *mapping*;
4. offrire strumenti utili alla verifica preliminare dei risultati.

Dopo una fase di *software selection*, basata su informazioni di pubblico dominio e in letteratura²⁹, si è identificata la suite D2RQ³⁰ sviluppata dalla Freie Universität Berlin come lo strumento più idoneo per la fase di trasformazione dei dati.

La suite si compone di vari strumenti dei quali sono stati utilizzati:

1. il linguaggio di *mapping* D2RQ-ML;
2. il server D2RQ-Server per l'implementazione di un *endpoint* SPARQL³¹;
3. l'estrattore Dump-RDF per la generazione del *dump* completo di *Corago*.

Nel complesso gli strumenti della suite si sono rivelati adeguati. In particolare, il linguaggio di *mapping* ha evidenziato una eccellente espressività unita ad una notevole semplicità di utilizzo. L'implementazione di costrutti di alto livello basati su una trasposizione dell'algebra relazionale permette di dichiarare i costrutti di *mapping* astraendosi dalla risoluzione delle query SQL necessarie alla produzione delle triple. Non si sono fatti approfondimenti sull'efficienza delle traduzioni, ma limitandosi alle esigenze di *dump*, il sistema non ha mai mostrato limiti significativi. Il server D2RQ-Server, utilizzato come supporto per visualizzare l'efficacia del *mapping* in corso di definizione, ha consentito di effettuare tutte le verifiche di congruità anche durante la fase di sviluppo. Quando si è passati alla successiva fase di verifica della qualità del dato, dovendo costruire query SPARQL più complesse, l'architettura che prevedeva l'accesso al file MDB via ODBC ha mostrato invece i suoi limiti non supportando alcuni costrutti SQL prodotti dalla libreria di trasformazione del server D2R. Si è quindi dovuto procedere all'individuazione di un server in grado di supportare sia la iniziale fase di verifica che di successiva consultazione del *dataset* ed è stato così scelto il *triple store* BlazeGraph³² prodotto da Systap.

²⁷ Un *endpoint* SPARQL è un servizio web che permette l'accesso ad un *dataset* RDF utilizzando il linguaggio SPARQL per la sua interrogazione.

²⁸ Estensione relativa al formato del file sorgente: Microsoft Access '97.

²⁹ In particolare, si è fatto riferimento a (Byrne, 2008), (Michel et. al., 2014) e (Bizer et al., 2009).

³⁰ Per una introduzione al sistema si rinvia a (Bizer and Seaborne, 2004). Per le specifiche funzionali e di sintassi del linguaggio D2RQ ML si rinvia al sito: <http://d2rq.org/> (verificato: 21/07/2019).

³¹ Si vedrà nel corso della trattazione come verranno allestite, sperimentate ed utilizzate diverse versioni di *endpoint* per il *dataset* utilizzando strumenti diversi in funzione della specifica fase del processo di trasposizione.

³² Sito web del produttore: <http://www.blazegraph.com/bigdata> (verificato: 21/07/2019).

Il diagramma 2.3 rappresenta l'architettura dei componenti utilizzati per la trasformazione ed estrazione:

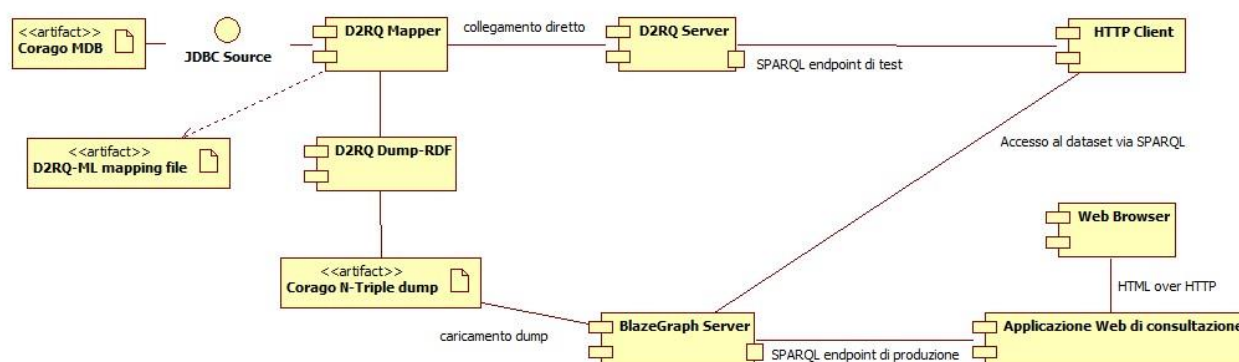


Diagramma 2.3: Architettura per componenti del sistema

2.2.3 Il mapping

È stata svolta una ricognizione su esperienze di utilizzo delle due ontologie di riferimento, su materiali e le linee guida prodotte dai gruppi di lavoro, su esperienze specifiche di trasposizioni di basi dati relazionali relative alle arti performative in RDF (Bellini P. e Nesi P., 2014) e (Byrne K., 2008), oltre che delle linee guida generali previste dal W3C³³.

Ne sono scaturite alcune considerazioni preliminari:

- a. in merito alla scelta del *pattern* per la formazione degli URI si è scelto di conservare l'indicazione dell'origine del dato in fonte (nome tabella e nome campo dove applicabile) unito alle chiavi relative alla proprietà definite come la chiave primaria del record di origine ed il valore dell'attributo³⁴;
- b. si è cercato di evitare la formazione di *blank node* arrivando con gli URI fino alle foglie con aggiunta di *label* per portare il contenuto informativo;
- c. nella esplicitazione dei tipi primitivi si è specificato il tipo fino alle proprietà *rdfs:label* o *rdfs:value*;
- d. in linea generale nella selezione delle classi su cui attestare le istanze si è scelto di seguire il principio di selezionare la classe *ancestor*³⁵ della classe FRBRoo se definita in CRM, in particolare in quei casi in cui la stessa viene definita come equivalente, seguendo un principio di massima generalizzazione senza perdita di significato;
- e. nell'implementare l'uso della classe E55_Type come *ancestor* astratto di tipizzazioni specifiche del dominio si è scelto di fare riferimento alla "tabella delle definizioni" TADE andando a specializzare E55_Type per ogni TADE_TIP. Solo dove la semantica del tipo in *Corago* è stata ritenuta fortemente qualificate si è proceduto alla definizione di una estensione di E55_Type in *Corago*;

³³ Vedi indirizzo: <http://www.w3.org/TR/rdb-direct-mapping/> (consultato il 21/07/2019).

³⁴ A titolo di esempio per Brani si ottiene un URI relativo: /BRANI/0000050272, per un URI derivato come Titolo: /TITOLI/BRANI&TITOLO_STP&0000050272. In questo caso l'URI risulta composto dal nome tabella fonte: TITOLI, dal nome tabella dell'entità associata: BRANI unita al TIPO di titolo e la chiave del Brano stesso (ovvero la FK relativa sulla tabella TITOLI).

³⁵ Il termine si riferisce alle classi che, in una gerarchia in cui viga il principio di ereditarietà, rappresentino una generalizzazione rispetto alla classe cui si fa riferimento. In senso assoluto è da intendersi come la classe archetipo dell'intera gerarchia.

- f. in vista dell'allineamento a LOD esterni per la georeferenziazione si è deciso di attestare Luoghi e Date sulle relative "Appellation" definite in CRM lasciando la relativa classe origine (ad es. F9_Place) associata alla classe di riferimento (ad es. F31_Performance) utilizzandone la relativa chiave primaria per generare l'URI;
- g. nell'aggiungere riferimenti ad ontologie esterne si è cercato di limitarsi a quelle strettamente necessarie ad esplicitare il significato del contenuto espresso in RDF; nonostante questo, data l'articolazione delle due ontologie di riferimento, l'inferenza di queste ha incrementato il numero di triple di un fattore 6 circa;
- h. per supportare le funzionalità di ricerca full-text delle entità si è provveduto a specificare una proprietà di tipo *rdfs:label* per tutte le entità che si vuole contemplare in questo tipo di ricerca³⁶.

Sulla base di queste linee guida si è proceduto con l'attività di analisi e sviluppo. La descrizione delle scelte fatte nella definizione del *mapping*, suddivise per le macro-aree tematiche è riportata dai diagrammi delle classi riportati nella appendice A. Data la dimensione complessiva, derivata anche dalla ripetitività di alcune aree, ne verrà data una trattazione sintetica, ma si evidenzieranno di volta in volta gli aspetti di natura concettuale più delicati o particolari soluzioni implementative adottate. Per gli aspetti strettamente tecnici si rimanda invece alla consultazione del file stesso di *mapping*³⁷ e per la relativa interpretazione alla documentazione del linguaggio D2RQ ML. Nel paragrafo 2.3, viene illustrata l'ontologia Corago SM in cui sono confluite le estensioni necessarie a colmare i limiti di espressività delle ontologie di riferimento rispetto alle specificità del nostro dominio.

³⁶ I *repository* utilizzati offrono soluzioni diverse per l'indicizzazione full-text, con funzionalità più o meno estese o configurabili, ma accomunate dalla necessità di decorare l'entità target con un testo sufficientemente qualificante ed indicativo dell'identità della stessa in modo da poter costruire un indice con un buon livello di discriminazione. Per questo, caso per caso, si è cercato di individuare l'attributo più saliente per ciascuna delle classi di entità che si sono mappate.

³⁷ Il file ed il progetto utilizzato per il *mapping* sono pubblicati attraverso la piattaforma GitHub all'indirizzo: <https://github.com/paolobonora/Corago-Mapping>.

2.2.4 Analisi della qualità del *dataset*

In questa fase si è cercato di verificare la corrispondenza dell'output del processo di estrazione dei contenuti ottenuta attraverso il *mapping* rispetto ai dati presenti nella fonte relazionale.

L'analisi è stata condotta su due livelli:

- la verifica della coerenza della semantica espressa attraverso le ontologie di riferimento rispetto a quella implicita del modello di origine;
- la verifica della qualità del dato esposto sia in termini di correttezza (corrispondenza tra dato in fonte e triple di destinazione) che di completezza (corrispondenza del conteggio delle entità presenti in fonte e nel *dataset* risultante).

La prima è una attività prevalentemente qualitativa, mentre la seconda si compone di analisi a campione e verifiche della copertura del *mapping* rispetto alla fonte.

Viste le dimensioni della fonte, l'eterogeneità dei contenuti e la necessità di competenze specifiche nei processi di catalogazione che hanno portato all'attuale dimensione della base dati d'origine, si è deciso di circoscrivere questa prima valutazione della qualità del dato ad un insieme di casi di test definiti con l'esperto di dominio.

I casi di test sono stati individuati per poter essere impiegati a due diversi livelli:

1. analisi quantitativa dell'output rispetto alla fonte;
2. analisi qualitativa della correttezza semantica del *mapping*.

Il primo livello ha come obiettivo la verifica che i criteri di estrazione adottati nel file di *mapping* abbiano mantenuto inalterato il rapporto tra istanze delle classi rappresentate e cardinalità delle righe nelle tabelle di origine. Questo criterio di verifica deve essere stringente per le classi il cui *mapping* è univoco rispetto alla fonte, ovvero l'URI dell'istanza è direttamente derivato dalla chiave primaria della tabella.

Per le classi derivate da associative (ad es. frbroo:E25_Performance_Plan da SPET_BRA) si è verificata la cardinalità rispetto alle chiavi esterne (*foreign keys* - FK) delle tabelle associate.

Per le classi, le cui istanze sono derivate da singoli campi di tabella (ad es. crm:E67_Birth), si è optato per una verifica a campione sui casi di test delle aggregazioni dei relativi valori (ad es. frbroo:F9_Place).

Dal punto di vista del processo relativo agli aspetti quantitativi si è proceduto come segue:

1. creazione del file di *dump* completo in formato N-Triple con la libreria D2RQ;
2. caricamento del *dump* in una istanza locale del *triple store* Blazegraph;
3. definizione delle query SPARQL per l'implementazione dei casi di test;
4. estrazione dei risultati e confronto con la fonte relazionale.

Per la valutazione degli aspetti qualitativi, si è verificato l'allineamento semantico dei due modelli ottenuti sia in fase di design che attraverso la lettura dei dati estratti per le diverse classi. Ad esempio, le responsabilità ed i relativi ruoli per le opere scelte come casi di test, oppure la coerenza tra cantanti e tipi di ruoli interpretati negli spettacoli. Questo tipo di verifica, oltre a considerare l'aderenza del dato semanticamente qualificato al dato in fonte, ha cercato anche di valutare la coerenza dello stesso rispetto al modello semantico su cui viene attestato in assenza di ogni riferimento con la fonte.

Di tipo intermedio sono le valutazioni che, col caricamento del *dump* sul server Blazegraph con calcolo delle inferenze, hanno permesso di evidenziare una anomalia nel *mapping* della relazione *frbroo:R25_performed*, per cui si era invertita in fase di *mapping* origine e destinazione col risultato di attribuire la classe *frbroo:F31_Performance* (definita come Domain della proprietà in FRBRoo) alle istanze di *frbroo:F25_Performance_Plan* (che ne rappresentano invece il Range) col risultato di avere un conteggio anomalo delle entità classificate come *frbroo:F31_Performance*.

Nel complesso questa prima fase di analisi ha cercato di validare l'output ottenuto dal *mapping* e di raffinarne la copertura del modello descrittivo (Console e Lenzerini, 2014).

2.2.5 Risultati

Raggiunto l'obiettivo primario della trasposizione dei contenuti e allestito un *endpoint* SPARQL si sono potuti valutare i risultati ottenuti in termini quantitativi e qualitativi relativi alle caratteristiche del *dataset* ottenuto.

2.2.5.1 Elementi quantitativi

Le misure quantitative restituiscono una prima sommaria indicazione circa l'estensione della copertura del modello di origine prodotta dal processo di trasposizione. Il numero di classi e di predicati sono indicativi dell'estensione dei concetti espressi, mentre il conteggio delle entità indicano la dimensione dei contenuti estratti³⁸.

Misura	Valore
Conteggio triple (senza inferenza) ³⁹	6.419.044
Conteggio triple (con inferenza) ⁴⁰	39.802.248
Conteggio classi distinte (senza inferenza)	43
Conteggio classi distinte (con inferenza)	95
Conteggio tipi di predicati (senza inferenza)	37
Conteggio tipi di predicati (con inferenza)	109

Tabella 2.2: Conteggi dimensioni del *dataset* RDF Corago

I risultati presentati in tabella 2.2 evidenziano l'effetto dell'inferenza sul numero di triple e sul significativo incremento del numero di classi che descrivono una risorsa. Inoltre, l'inferenza delle ontologie di riferimento comporta una significativa estensione della classificazione.

Una prima conseguenza è la possibilità di esporre interfacce con diversa granularità semantica verso l'esterno. Ad esempio, un sistema esterno potrà accedere alle risorse Corago utilizzando indifferentemente uno qualsiasi dei livelli di specializzazione dei concetti definiti nelle ontologie inferite. Ciascuna istanza di risorsa Corago potrà essere 'vista' sia come *crm:E1_CRM_Entity* che come membro della classe asserita come

³⁸ I dati riportati nella tabella 2.2 si riferiscono alla consistenza del dataset al 24/03/2018.

³⁹ Indica le classi delle istanze generate direttamente come output del *mapping* in fase di *dump (stated)*.

⁴⁰ Indica il numero complessivo delle classi originate dal *mapping (stated)* e di quelle inferite in base alle ontologie utilizzate.

assioma (*stated*) utilizzata come target nel *mapping* (ad. Es. frbroo:F31_Performance), oltre a tutte le classi intermedie della gerarchia definita in CRM e FRBRoo. La disponibilità di diversi livelli di astrazione, se da un lato rende più semplice l'interoperabilità verso l'esterno, dall'altro pone la necessità di individuare livelli omogenei di classificazione. In assenza di questi si determinano evidenti asimmetrie nella granularità della definizione semantica degli oggetti descritti (più alta è la generalizzazione, minore è il potere discriminante della tassonomia).

Una seconda conseguenza della significativa estensione della classificazione dovuta all'inferenza è la necessità di identificare il livello nativo o più fine della classificazione delle risorse. Qualora si ponga la necessità di attestare caratteristiche specifiche ad una risorsa in funzione della sua classe *stated*, diventa necessario ottenere per ciascun oggetto il suo tipo 'nativo' o comunque un particolare livello di generalizzazione.

2.2.5.2 Elementi qualitativi

Una volta definito il perimetro oggetto dell'intervento di trasposizione, resta da determinare l'eventuale completezza della rappresentazione prodotta. Si è cercato di rappresentare almeno tutti gli attributi identitari di ciascuna entità (ad es. tutte chiavi naturali delle tabelle di origine) e di quegli attributi più qualificanti per il dominio. Per una valutazione oggettiva si dovrebbe costruire una misura che identifichi la copertura dei campi estratti rispetto ai campi presenti nel modello fisico del DBMS. Resterebbe comunque da valutare la reale rilevanza delle esclusioni fatte. Un diverso approccio potrebbe essere basato su una analisi di "autosufficienza" della rappresentazione ottenuta. Ad esempio, la capacità della rappresentazione di supportare analisi dei duplicati di entità⁴¹, o la possibilità di inferire relazioni con *dataset* LOD esterni sulla base delle caratteristiche delle reciproche entità equivalenti.

⁴¹ In questo caso si sta facendo qualche esperimento per costruire query SPARQL per la deduplicazione dei 'Responsabili'. In questa applicazione diventa rilevante l'esposizione di tutti gli attributi identitari nella corretta rappresentazione semantica. I risultati ottenuti si potrebbero confrontare con le tecniche applicabili direttamente sulla base dati relazionale.

2.3 Specializzazione del modello concettuale: il Corago Semantic Model

L'ontologia *Corago Semantic Model* (Corago SM)⁴² è stata progettata per essere una naturale estensione del CIDOC CRM e di FRBRoo e quindi risulta nativamente allineata. La sua funzione principale è integrare e specializzare alcune definizioni di concetti definiti dalle due ontologie di riferimento per poter rappresentare in modo più aderente alle esigenze di dominio alcuni contenuti dell'archivio. Le estensioni hanno riguardato l'introduzione di nuove classi, di una serie di *object properties* ed alcune *data properties*.

L'ontologia è definita mediante il linguaggio OWL-DL 1.0 ed include nella sua definizione l'importazione dei tipi definiti in CIDOC CRM e FRBRoo nella cosiddetta versione "Erlangen"⁴³.

Il diagramma sottostante rappresenta le classi introdotte e la relativa dipendenza nelle ontologie di riferimento.

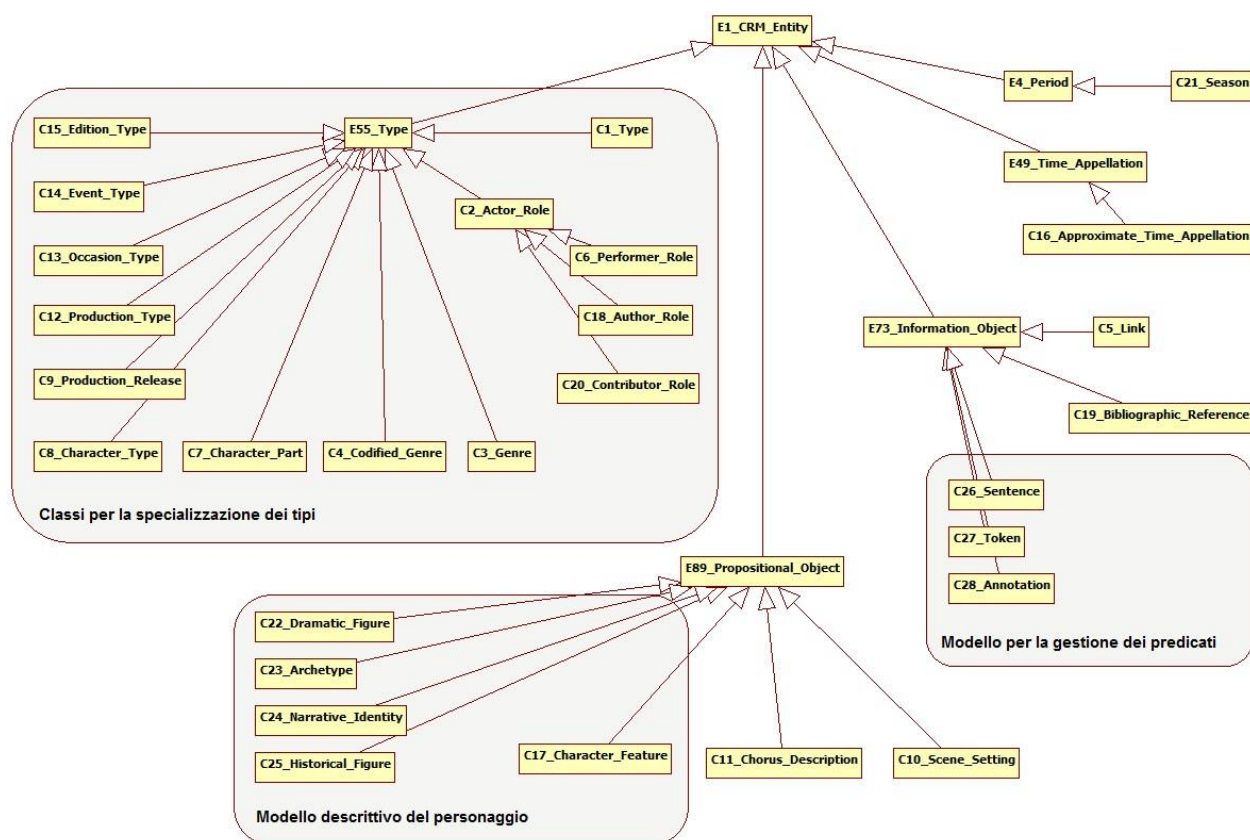


Diagramma 2.4: Classi dell'ontologia Corago SM

Nella tabella 4 vengono riportate le classi principali e le relative proprietà del Corago SM. Per la definizione analitica della gerarchia delle dipendenze si rinvia alla versione integrale dell'ontologia accessibile all'indirizzo: <https://github.com/paolobonora/Corago-SM>.

⁴² L'ontologia è pubblicata attraverso il *repository* GitHub all'indirizzo: <https://github.com/paolobonora/Corago-SM>.

⁴³ La versione Erlangen CRM/OWL è una implementazione in linguaggio OWL-DL 1.0 del CIDOC CRM realizzata e mantenuta dall'Università Friedrich Alexander di Erlangen-Nuremberg. Recentemente è stata pubblicata anche per FRBRoo una versione nello stesso formato.

Classe	Descrizione	URI
C2 Actor Role	Defines or specifies the role that an Actor has had in a specific Activity. Should be used with CP2 and CP3 in order to link an E7_Activity to its E39_Actor with a specific Role. Could be considered the long path of P14_carried_out_by.	http://corago.unibo.it/sm/C2_Actor_Role
C3 Genre	Defines the Genre of the E1_CRM_Entity.	http://corago.unibo.it/sm/C3_Genre
C4 Codified Genre	Defines the Codified Genre of the E1_CRM_Entity.	http://corago.unibo.it/sm/C4_Codified_Genre
C5 Link	Defines the type of link between F1_Work or F2_Expression.	http://corago.unibo.it/sm/C5_Link
C6 Performer Role	Defines or specifies the role with a Character or Performance has to be performed with.	http://corago.unibo.it/sm/C6_Performer_Role
C7 Character Part	Defines the kind of part taken by the character in the Performance.	http://corago.unibo.it/sm/C7_Character_Part
C8 Character Type	Describes the stereotyped kind of Character.	http://corago.unibo.it/sm/C8_Character_Type
C9 Production Release	Defines the kind of production release in order of its public presentation as premiere or a reprise.	http://corago.unibo.it/sm/C9_Production_Release
C10_Scene_Setting	Describes the scene setting of the work	http://corago.unibo.it/sm/C10_Scene_Setting
C11_Chorus_Description	Describes the chorus characteristics, its role or part in the work and composition	http://corago.unibo.it/sm/C11_Chorus_Description
C12_Production_Type	Defines the production type	http://corago.unibo.it/sm/C12_Production_Type
C13_Occasion_Type	Describes the kind of the occasion being celebrated with the Performance	http://corago.unibo.it/sm/C13_Occasion_Type
C14_Event_Type	Describes the type or kind of the event being performed	http://corago.unibo.it/sm/C14_Event_Type
C15_Edition_Type	Defines the type of edition of a document (i.e. first edition, reprint, etc.)	http://corago.unibo.it/sm/C15_Edition_Type
C16_Approximate_Time_Appellation	Represents the best approximate definition available for a Time Span	http://corago.unibo.it/sm/C16_Approximate_Time_Appellation
C17_Character_Feature	This class represents a description of a character's atomic feature as derived from the character's description found in sources. It may represent both physical or abstract characteristic of the related character	http://corago.unibo.it/sm/C17_Character_Feature
C18_Author_Role	Defines a role of authorship carried out by an Actor such as work's conception, expression creation but not role acting	http://corago.unibo.it/sm/C18_Author_Role
C19_Bibliographic_Reference	Defines a direct reference between two E31_Document instances. It is used to refer to a bibliographic reference	http://corago.unibo.it/sm/C19_Bibliographic_Reference
C20_Contributor_Role	This class defines a generic contribution role of the Actor differing from authorship or performing	http://corago.unibo.it/sm/C20_Contributor_Role
C21_Season	Identifies a Period of time in a year while a set of events are performed	http://corago.unibo.it/sm/C21_Season
C22_Dramatic_Figure	Representative figure of one or more homologous characters, that is superimposable in terms of characteristics and role played in their respective dramatic actions appearing in different works	http://corago.unibo.it/sm/C22_Dramatic_Figure
C23_Archetype	A set of attributes characterizing a type of Dramatic Figure	http://corago.unibo.it/sm/C23_Archetype
C24_Narrative_Identity	A subject with its own identity and a space-time collocation that is represented in moments of its existence or with attributes different from one or more dramatic figures that are referred to as a unique and autonomous entity	http://corago.unibo.it/sm/C24_Narrative_Identity
C25_Historical_Figure	Historically attested actor or entity to which a narrative identity or an archetype refers to	http://corago.unibo.it/sm/C25_Historical_Figure

C26_Sentence	A sentence expressed in natural language	http://corago.unibo.it/sm/C26_Sentence
C27_Token	A Part Of Speech token	http://corago.unibo.it/sm/C27_Token
C28_Annotation	A morphosyntactic annotation resulting from a NLP analysis of a text	http://corago.unibo.it/sm/C28_Annotation
Proprietà	Descrizione	URI
CP2 carried out role	This property defines the relation between an E7 Activity and a C2 Actor Role which is the role an E39 Actor has carried it out with.	http://corago.unibo.it/sm/CP2_carried_out_role
CP3 carried out actor	This property defines the relation between a C2 Actor Role for a specific E7 Activity and the E39 Actor who has carried it out.	http://corago.unibo.it/sm/CP3_carried_out_actor
CP4 links ⁴⁴	Represent a direct (untyped) relation between two entities. Equivalent to P67 refers to, could be inferred by the property chain: CP5 refers with o CP6 refers to.	http://corago.unibo.it/sm/CP4_links
CP5 refers with	Relates the domain range entity with the instance of C5 Link which specifies the nature of the relation occurring with the target entity it is referring to through CP6 refers to.	http://corago.unibo.it/sm/CP5_refers_with
CP6 refers to	Relates the C5 Link which specifies the nature of the relation occurring with the range entity it is referring to.	http://corago.unibo.it/sm/CP6_refers_to
CP7 performed part	Relates the C2 Actor Role with the part performed by the actor in the role.	http://corago.unibo.it/sm/CP7_performed_part
CP8 performed character	Specifies the Character performed with the Actor Role.	http://corago.unibo.it/sm/CP8_performed_character
CP9_first_performed_in	Identifies the first performance of a work	http://corago.unibo.it/sm/CP9_first_performed_in
CP10_documented_in	Documents the source of an Actor role	http://corago.unibo.it/sm/CP10_documented_in
CP11_first_performance_place	Place where the work has been performed first	http://corago.unibo.it/sm/CP11_first_performance_place
CP12_first_performance_location	Location where the work has been performed first	http://corago.unibo.it/sm/CP12_first_performance_location
CP13_first_performance_date	First performance date	http://corago.unibo.it/sm/CP13_first_performance_date
CP14_conception_date	Date of conception of a work	http://corago.unibo.it/sm/CP14_conception_date
CP15_has_URL	This property specifies a navigable URL for a resource	http://corago.unibo.it/sm/CP15_has_URL
CP16_book_title	Specifies the title of a book as stated within the source	http://corago.unibo.it/sm/CP16_book_title
CP17_most_specific_title	Specifies the most specific title of a book as stated within the source	http://corago.unibo.it/sm/CP17_most_specific_title
CP18_has_parent_figure	Specifies a dramatic feature's parent	http://corago.unibo.it/sm/CP18_has_parent_figure
CDP1_work_genre_in_source	Genre in source of a work	http://corago.unibo.it/sm/CDP1_work_genre_in_source
CDP2_is_attribution_uncertain	If true, the Actor Role is uncertain attributed to the Actor	http://corago.unibo.it/sm/CDP2_is_attribution_uncertain
CDP3_is_secondary_source_attribution	True when the attribution has been made upon a secondary or complementary source	http://corago.unibo.it/sm/CDP3_is_secondary_source_attribution
CDP4_is_primary	Identifies the primary authorship role	http://corago.unibo.it/sm/CDP4_is_primary
CDP5_season_sequence_order	Identifies the ordinal position of the performance in a Season events sequence	http://corago.unibo.it/sm/CDP5_season_sequence_order
CDP6_performance_title_in_source	The title of the performance as reported by the source	http://corago.unibo.it/sm/CDP6_performance_title_in_source

⁴⁴ La proprietà viene dichiarata come equivalente a `crm:P67_refers_to` ed inferibile in base alla catena di proprietà: `corago:CP5_refers_with` e `corago:CP6_refers_to` attraverso la classe `corago:C5_Link`.

CDP7_document_editorial_number	Editorial number of a document	http://corago.unibo.it/sm/CDP7_document_editorial_number
CDP8_document_edition_in_source	Edition of a document as reported by the source	http://corago.unibo.it/sm/CDP8_document_edition_in_source
CDP9_work_title_in_source	The title of the work as reported by the source	http://corago.unibo.it/sm/CDP9_work_title_in_source
CDP10_bibliography_author_in_source	Autor of the bibliographic reference reported by the source	http://corago.unibo.it/sm/CDP10_bibliography_author_in_source
CDP11_document_title_in_source	Represents the title as reported by the source	http://corago.unibo.it/sm/CDP11_document_title_in_source
CDP12_has_digital_reproduction	URL of the digital reproduction of the resource	http://corago.unibo.it/sm/CDP12_has_digital_reproduction
CDP13_location	Represents the location of the referred entity in the lat-lon format (http://www.bigdata.com/rdf/geospatial/literals/v1#lat-lon) required to perform geospatial query. Example: "48.13743#11.57549"^^< http://www.bigdata.com/rdf/geospatial/literals/v1#lat-lon >	http://corago.unibo.it/sm/CDP13_location
CDP14_happened	Represent the point in space and time when the referred entity occurred expressed in the lat-lon-time format (http://www.bigdata.com/rdf/geospatial/literals/v1#lat-lon-time). Example: "48.13188#11.54965#1379714400"^^< http://www.bigdata.com/rdf/geospatial/literals/v1#lat-lon-time >	http://corago.unibo.it/sm/CDP14_happened
CDP15_location_from_scene_setting	Location derived from the scene setting of a work	http://corago.unibo.it/sm/CDP15_location_from_scene_setting
CDP16_location_from_description	Location derived from a description of a work or a character setting of a work	http://corago.unibo.it/sm/CDP16_location_from_description
CDP17_location_from_performance	Location derived from a performance	http://corago.unibo.it/sm/CDP17_location_from_performance
CDP18_is_literary_text_original	Specifies if the literary text used in the libretto is original (Boolean value)	http://corago.unibo.it/sm/CDP18_is_literary_text_original
CDP19_literary_text_originality	Specifies if the literary text of the libretto is original or a reprise or a variation of (textual value)	http://corago.unibo.it/sm/CDP19_literary_text_originality

Tabella 2.3: Classi e proprietà dell'ontologia Corago SM

2.3.1 Rappresentazione delle responsabilità autoriali

Se nel complesso le due ontologie scelte come riferimento per rappresentare i contenuti del dominio si sono rivelate efficaci. Le specializzazioni aggiunte sono state necessarie soltanto per non perdere il significato settoriale di alcune informazioni come nel caso della rappresentazione del ruolo svolto da un autore od interprete si è rivelato necessario introdurre una significativa estensione del modello previsto a questo scopo. Il CIDOC-CRM definisce la proprietà `crm:P14_carried_out_by` per associare l'autore (`cm:E39_Actor`) all'attività di cui è 'responsabile' e una "sub-proprietà" `crm:P14.1_in_the_role_of` per specificare in un'istanza `crm:E55_Type` il ruolo nell'azione. Questo schema, sebbene concettualmente corretto, non è direttamente implementabile in RDF. Ciò richiede l'adozione di alcune soluzioni tecnicamente fattibili ma che necessariamente introducono lunghe catene di proprietà tra le due entità principali coinvolte nella relazione di 'attribuzione della responsabilità autoriale' (Alexiev, 2012).

D'altra parte, nel teatro d'opera il legame tra autore ed opera non può prescindere dalla possibilità di specificarne il ruolo svolto nella sua concezione. Ad esempio, non ha senso indicare Giuseppe Verdi e Francesco Maria Piave come autori de *La traviata* senza specificare che il primo ne è il compositore delle musiche ed il secondo l'autore del libretto.

Per sopperire a questo limite si è deciso di estendere il modello previsto dal CIDOC CRM introducendo la classe corago:C2_Actor_Role come classe intermedia per qualificare la relazione di autorialità.

Ad esempio per rappresentare la relazione tra un'opera (definita in base alla classe `crm:F1_Work`) e i suoi autori si deve prevedere la definizione di una istanza della classe `frbroo:F27_Work_Conception` che rappresenta l'attività relativa alla sua concezione. Questa sarà poi associata con una `R16_was_initiated_by` all'opera che ha prodotto e con una proprietà `corago:CP2_carried_out_role` ad una istanza `corago:C2_Actor_Role` a sua volta associata attraverso una `corago:CP3_carried_out_actor` alle relative istanze di `crm:E39_Actor`.

L'istanza della classe `corago:C2_Actor_Role` permette di specificare, attraverso i suoi attributi, il "ruolo" svolto nel processo di concezione dell'opera. Diventa in questo modo possibile specificare i rispettivi ruoli di Verdi e Piave nel realizzare la *La traviata*. La stessa struttura può essere applicata in tutti i casi in cui è prevista la possibilità di associare due classi attraverso la proprietà `crm:P14_carried_out_by`⁴⁵, ad esempio nel caso del ruolo svolto da un interprete in uno spettacolo.

L'unico svantaggio di questa soluzione è l'introduzione di una catena di proprietà (*property chain*) al posto di una relazione espressa da una proprietà semplice. L'immediata conseguenza è l'aumento della complessità del grafo risultante che, in questo caso, risulta ampiamente giustificata se commisurata alla perdita di espressività rispetto ai contenuti dell'archivio.

Tra le diverse estensioni del modello di riferimento, quest'ultima è forse quella che caratterizza maggiormente l'intero processo di trasposizione. Non solo perché evidenzia la necessità di una estensione del modello concettuale di riferimento in relazione al dominio di applicazione, ma anche perché la gestione della complessità che questa introduce sarà uno dei fattori che giustificano l'introduzione dell'ulteriore livello descrittivo orientato alla navigazione dei contenuti descritto nel paragrafo 2.5.

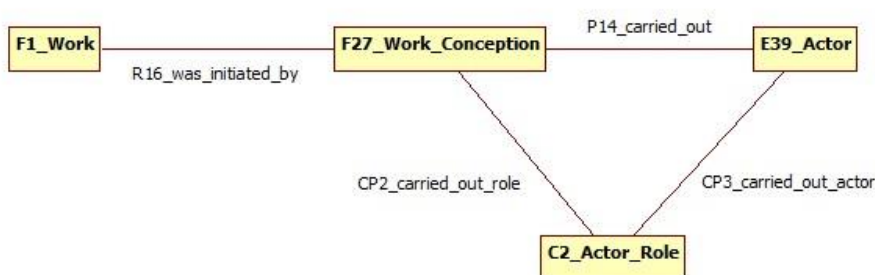


Figura 2-1: Relazione di responsabilità

⁴⁵ Pur non essendo definita formalmente come tale, dal punto di vista concettuale, la catena costituita dalla proprietà `corago:CP2_carried_out_role` associata ad una `corago:C2_Actor_Role` ed alla `corago:CP3_carried_out_actor` è concettualmente una sotto-proprietà equivalente della `crm:P14_carried_out_by`.

2.3.2 Rappresentazione del ruolo dell'interprete

Un ulteriore caso che ha richiesto una estensione del modello descrittivo offerto dalle due ontologie di riferimento è dato dalla relazione il personaggio ed il suo interprete. Così come per gli autori delle opere, anche la relazione tra il cantante e la parte è descritta in modo analitico dal modello concettuale Corago. Gli attributi che la specificano vanno dal ruolo al nome dell'interprete ricavato dalla fonte (che può differire da quello anagrafico). Per raccogliere l'insieme delle caratteristiche della relazione è stata introdotta la classe corago:C6_Performer_Role. Nata come specializzazione della classe corago:C2_Actor_Role descrive il ruolo specifico dell'interprete in uno spettacolo associandolo direttamente al relativo frbroo:F25_Performance_Plan tramite la proprietà corago:CP7_performed_part ed al personaggio interpretato attraverso la proprietà corago:CP8_performed_character.

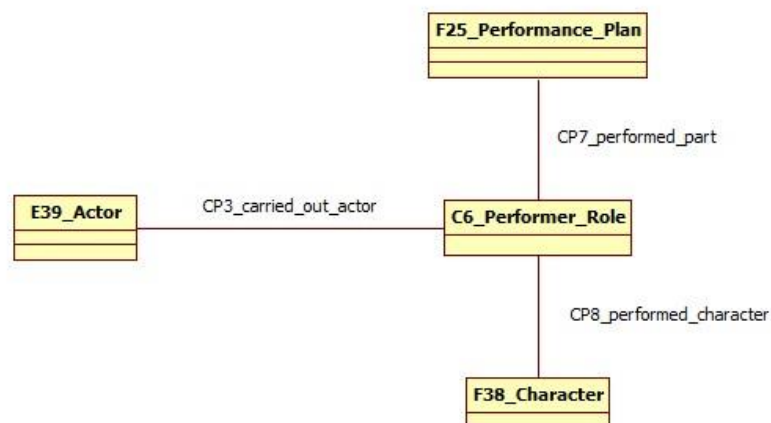


Figura 2-2: Relazione tra interprete, spettacolo e personaggio

Anche in questo caso il modello proposto permette di rappresentare compiutamente l'articolazione della relazione e l'insieme di attributi associati pur rimanendo coerente col modello generale del CIDOC CRM.

2.4 L'organizzazione dei contenuti nella prospettiva funzionale FRBR

Il modello concettuale FRBR si è posto l'obiettivo di strutturare i contenuti dell'informazione bibliografica in modo da rispondere efficacemente alle esigenze di consultazione degli utenti finali. L'articolazione gerarchica del modello FRBR, procedendo dal piano astratto dell'opera fino al piano materiale dell'esemplare, offre la possibilità di scegliere il punto di ingresso da cui avviare l'esplorazione dei contenuti dell'archivio. Attraverso i rinvii tra le entità è possibile muoversi tra i diversi livelli: il riferimento bibliografico di un'opera permette di spostarsi dal contenuto (l'opera), al contenitore (il volume che la contiene).

Nell'archivio *Corago* questa gerarchia si sviluppa a partire dal Repertorio delle opere e trova sua espressione nella Cronologia degli spettacoli. I Libretti, a loro volta, sono la testimonianza materiale dello Spettacolo e che la fonte documentale principale del testo. Ciascuna delle entità presenti nei tre livelli viene descritta attraverso una scheda che ne individua le caratteristiche principali e ne specifica le relazioni con le altre entità del dominio. Gli attributi che compongono la scheda hanno diversa rilevanza e possono assumere valenze particolari al variare del contesto funzionale.

Si è già evidenziato nel paragrafo 1.5 come questo diverso valore funzionale dell'attributo introduca tre specifici requisiti con l'obiettivo comune di produrre una rappresentazione dei contenuti efficace rispetto ai due principali scenari d'uso: la ricerca e la consultazione dei contenuti dell'archivio.

In particolare, il requisito [REQ1] evidenzia la necessità di poter specificare la rilevanza di un particolare attributo come chiave per determinare l'identità delle istanze di ciascuna tipologia di entità. Questo lo qualifica come valido candidato a rientrare tra i campi da inserire nelle funzioni di ricerca per quella specifica tipologia di entità (ovvero classe nel modello *object oriented* dell'ontologia di dominio). Questa scelta può avvenire in modo implicito, ovvero inserendo il campo nelle maschere dell'applicazione di consultazione dell'archivio, oppure in modo esplicito introducendo un livello descrittivo nel modello concettuale attraverso cui indicare quali, tra gli attributi che descrivono una classe di entità, debbano essere utilizzati come criteri di ricerca.

Parallelamente, i requisiti [REQ2] e [REQ3] evidenziano come alcuni attributi assumano particolare efficacia nel rappresentare le caratteristiche degli individui di una classe.

In particolare, il [REQ2] pone il problema di come specificare quali siano gli attributi attraverso cui l'utente è in grado di stabilire l'identità delle istanze di una classe del dominio. Ad esempio, titolo, autori e anno di prima esecuzione sono gli elementi minimi per capire di che opera si tratti; così come per uno spettacolo è imprescindibile conoscere almeno data e luogo della rappresentazione.

L'introduzione di questo livello di annotazione all'interno del modello concettuale diventa necessaria anche per definire formalmente un criterio di confronto utile a stabilire se due istanze debbano essere considerate sovrapponibili o equivalenti. Gli attributi che vengono utilizzati in questa definizione possono essere classificati come 'identitari' per una tipologia di entità.

Attraverso l'introduzione di una gerarchia tra gli attributi 'identitari' di una classe, diventa possibile stabilire un criterio di misura della distanza (o similarità) tra le sue istanze. Ad esempio, se tre rappresentazioni condividono lo stesso programma, ma solo due hanno in comune parte della compagnia di interpreti, mentre nella terza questa differisce totalmente, si potrebbe stabilire che, pur trattandosi di tre spettacoli diversi, dei tre è il terzo che "differisce maggiormente"⁴⁶.

⁴⁶ Si veda nel capitolo 3 come il confronto di similarità basato su specifici attributi descrittivi possa essere applicato per stabilire relazioni tra i testi letterari.

In termini funzionali, la necessità di confrontare le diverse istanze in base ad un sottoinsieme di loro caratteristiche risulta evidente nella consultazione degli elenchi di risultati delle ricerche. In questo caso è necessario presentare all'utente un insieme coerente e costante di attributi 'identitari' per ciascuna tipologia di risultato in modo da consentirgli di discriminare facilmente le diverse occorrenze.

In questo contesto, la stessa modalità di presentazione è rilevante in termini di usabilità. Il modello di annotazione degli attributi deve quindi prevedere la possibilità di indicare anche gli elementi utili a realizzare rappresentazioni efficaci dell'attributo. In questo senso il requisito [REQ3] richiede di introdurre nel modello annotazioni utili a gestire le modalità di presentazione di ciascun attributo nei diversi contesti d'uso.

L'oggetto di questo livello di annotazione degli attributi deve quindi rispondere sia a requisiti relativi alla 'rilevanza semantica' dell'attributo, sia di 'rappresentazione formale' in prospettiva funzionale. L'approccio che è stato adottato e sperimentato prevede la derivazione degli elementi utili a produrre una rappresentazione delle singole entità a partire dalla rilevanza semantica dell'attributo all'interno del rispettivo modello descrittivo.

Ad esempio, il fatto che in base a FRBR sia sufficiente il solo titolo a qualificare un'opera, fa sì che questo sia contemporaneamente il primo attributo 'identitario' della classe e l'elemento informativo più efficace per rappresentarla. Partendo da questa considerazione, l'ipotesi su cui si è lavorato prevede che, una volta stabilita una gerarchia tra gli attributi descrittivi della classe in grado di rendere possibili confronti diretti tra istanze, la stessa gerarchia possa essere utilizzata per costruire rappresentazioni efficaci anche in prospettiva utente.

Questo approccio, che affianca al modello concettuale definito in base alle due ontologie di riferimento un ulteriore livello descrittivo della rilevanza semantica dei contenuti in prospettiva funzionale, ha spinto alla definizione del paradigma di navigazione basato sulla rappresentazione formale della semantica di dominio presentato nel paragrafo 2.5.

2.5 Navigazione guidata dalla semantica di dominio

2.5.1 Introduzione

L'impiego delle due ontologie CIDOC CRM e FRBRoo utilizzate per rappresentare i contenuti dell'archivio ha determinato la complessa morfologia del grafo risultante. Se da un lato l'organizzazione dell'informazione che si è ottenuta risponde ai criteri di coerenza e generalità delle due ontologie di riferimento, dall'altro questa articolazione costituisce l'ineludibile punto di partenza per l'implementazione delle funzionalità di consultazione.

La complessità introdotta a valle della trasposizione richiede che il sistema di consultazione operi una forma di mediazione tra utente e dato. Sia in fase di interrogazione che di rappresentazione dei contenuti. Si tratta quindi di individuare una soluzione che permetta di definire le modalità con cui le funzionalità di consultazione dovranno gestire questa interazione.

Nell'ambito delle tecnologie semantiche sono state proposte diverse soluzioni per la realizzazione di viste utente che vanno dalla creazione di interrogazioni dirette del grafo all'introduzione delle cosiddette 'lenti semantiche' (Colazzo et al., 2014). Il limite principale di questo tipo di soluzione sta nel produrre elementi definiti al di fuori del perimetro del modello di dominio e disegnati come strumenti finalizzati alla sola implementazione dei requisiti di accesso all'informazione. Si tratta di strumenti per la realizzazione di viste applicative che operano sul dato e ne producono una rappresentazione per un uso specifico. Le logiche in base alle quali il dato viene selezionato e presentato restano circoscritte all'implementazione della vista all'interno dell'applicazione di consultazione.

L'approccio seguito in questa sperimentazione richiede che venga prodotta una definizione formale degli elementi in base a cui verranno derivate le viste applicative. Ovvero, nel nostro modello, la rappresentazione dell'informazione è direttamente derivata dal valore semantico che il dato assume all'interno del dominio. La rilevanza di un attributo caratteristico di una classe determina pertanto la modalità con cui dovrà essere costruita una sua efficace rappresentazione. Al tempo stesso sarà necessario mantenere su due livelli distinti la definizione formale della semantica di dominio dalla modalità di rappresentazione dei contenuti che verrà quindi implementata a partire dalla prima.

La soluzione individuata prevede l'inserimento nel modello concettuale di componenti che individuino le caratteristiche rilevanti delle diverse entità del dominio ai fini della loro rappresentazione, ricerca ed indicizzazione. Questi componenti costituiscono un livello di astrazione intermedio tra il dato definito sulla base delle ontologie di riferimento e la sua rappresentazione destinata all'utente. La loro definizione in OWL-DL all'interno dell'ontologia Corago SM fa sì che entrino a far parte della base di conoscenza.

2.5.2 Entità ed attributi: dalla morfologia del grafo alla rappresentazione funzionale

Le due ontologie utilizzate per la definizione del *dataset* sono ontologie formali che presentano un notevole livello di analiticità nella definizione dei concetti con particolare attenzione alle dinamiche temporali, quindi di trasformazione nel tempo delle entità descritte. La granularità fine delle definizioni è motivata dalla riduzione dell'alone semantico del concetto rappresentato con la diretta conseguenza di ridurne l'ambiguità rendendolo quindi applicabile in contesti diversi.

Ne derivano una precisione e una generalità del modello descrittivo risultante, qui intesa come minore specializzazione dei singoli concetti di base ad uno specifico dominio. Per contro, nel momento in cui è necessario esprimere concetti complessi legati ad uno specifico dominio, risulta necessario comporre più

concetti elementari attraverso i nessi logici, ovvero i predicati o relazioni, messi a disposizione dal modello concettuale dell'ontologia. Si riportano due esempi della complessità risultante dalla composizione necessaria a descrivere due relazioni caratteristiche del nostro contesto di riferimento: quella tra un'opera e gli spettacoli in cui viene messa in scena e quella tra il libretto ed il suo editore.

Nel primo caso, l'opera e lo spettacolo sono rappresentate rispettivamente dalle classi `frbroo:F1_Work` e `frbroo:F31_Performance`. La relazione tra le due è mediata dalla `frbroo:F25_Performance_Plan` che rappresenta a livello di *expression* il contenuto dello spettacolo. Le classi `frbroo:F1_Work` e `frbroo:F25_Performance_Plan` sono collegate dalla proprietà `frbroo:R9_is_realised_in` che rappresenta la relazione tra opera ed espressione, mentre l'indicazione di quali contenuti sono stati rappresentati in uno spettacolo è espressa mediante la proprietà `frbroo:R25_performed` che associa una `frbroo:F31_Performance` al relativo `frbroo:F25_Performance_Plan`.

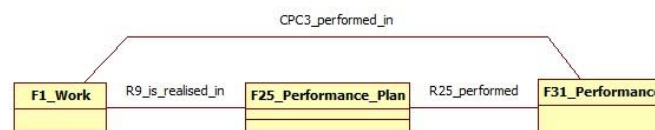


Diagramma 2.5: Relazione tra opera e spettacolo

Nel secondo caso, la relazione tra libretto ed il suo editore risulta ancora più articolata. E' necessario infatti collegare la classe `frbroo:F3_Manifestation_Product_Type`, utilizzata per rappresentare il libretto, con la classe `crm:E39_Actor` che rappresenta l'editore. In FRBRoo il nesso è l'evento di pubblicazione dei contenuti del libretto, rappresentato dalla classe `frbroo:F30_Publication_Event`, a loro volta rappresentati dalla classe `frbroo:F24_Publication_Expression`. L'informazione relativa all'editore di un libretto viene quindi rappresentata dalla concatenazione delle relazioni: `frbroo:CLR6_should_carry` che associa il libretto (`frbroo:F3_Manifestation_Product_Type`) ai suoi contenuti (`frbroo:F24_Publication_Expression`); la relazione tra questi e l'evento di pubblicazione (`frbroo:F30_Publication_Event`), definita dalla proprietà `crm:P94_has_created` ed infine la relazione tra l'evento e l'editore (`crm:E39_Actor`) mediante la proprietà `crm:P14_carried_out_by`. Per completezza sarebbe necessario considerare anche la specifica denominazione dell'editore derivata dalla fonte che è rappresentata mediante la classe `crm:E82_ActorAppellation` associata mediante la proprietà `crm:P131_is_identified_by`.

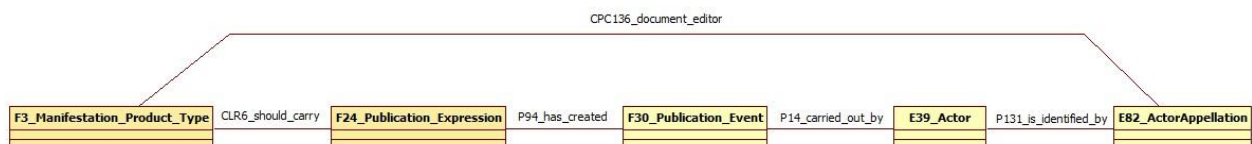


Diagramma 2.6: Relazione tra libretto ed editore

In entrambi i casi, l'articolazione prevista dai modelli concettuali delle due ontologie mira a definire senza ambiguità le entità coinvolte nella rappresentazione dei fatti e le relazioni che intercorrono tra esse. La complessità deriva dall'esigenza di esprimere, in termini coerenti col modello FRBR, il passaggio dal piano concettuale (espresso mediante le classi `frbroo:F1_Work` e `frbroo:F3_Manifestation_Product_Type`), al piano dei contenuti (concetti espressi con le classi `frbroo:F25_Performance_Plan` e

frbroo:F24_Publication_Expression) alle attività che ne rendono manifesta l'espressione (classi frbroo:F31_Performance e frbroo:F30_Publication_Event).

Per contro, dal punto di vista funzionale, le due relazioni descritte risultano facilmente sintetizzabili in linguaggio naturale come: “gli spettacoli in cui è stata rappresentata un’opera” e “l’editore del libretto”. Questo è il livello descrittivo richiesto dell’utente e che dovrà essere gestito nell’ontologia di dominio in modo coerente al modello concettuale sottostante. A questo scopo diventa quindi necessario poter definire le due relazioni in forma ‘sintetica’ a partire dalla catena di relazioni ‘analitiche’ sottostanti. Le relazioni sintetiche saranno il risultato della composizione del grafo di entità e proprietà che analiticamente descrivono i fatti all’interno del *dataset*. Queste relazioni erediteranno dalla catena sottostante la cardinalità e dovranno rappresentare in forma sintetica la semantica espressa dalle singole componenti.

Le informazioni possono essere così rappresentate attraverso due piani sovrapposti. Alla base i contenuti espressi dal modello concettuale delle ontologie di riferimento e, ad un livello superiore, la rappresentazione degli stessi secondo la prospettiva funzionale del dominio.

2.5.3 Modello delle proprietà di navigazione

L’implementazione dello strato intermedio necessario all’implementazione della soluzione proposta sfrutta le potenzialità offerte dal linguaggio OWL-DL nel definire le proprietà di tipo oggetto (owl:ObjectProperty) come la composizione di una catena di altre proprietà (*property chain*) indicandone la concatenazione attraverso il costrutto owl:propertyChainAxiom. Le proprietà così definite sono, dal punto di vista semantico, del tutto equivalenti alle relative *property chain* e costituiscono una sintesi formale della relazione tra le entità che mettono in relazione ‘diretta’. Al tempo stesso risultano elementi sovraordinati alle *property chain* sottostanti: si realizza così uno strato distinto dal livello rappresentato dalle ontologie di riferimento attraverso cui modellare gli aspetti specifici del dominio.

Le proprietà con queste caratteristiche, che assolvono al ruolo funzionale di guidare la navigazione della conoscenza del dominio, possono essere etichettate come ‘proprietà di navigazione’. L’archetipo (*ancestor*) comune di tutte queste proprietà è definito nell’ontologia di dominio dalla proprietà astratta corago:CPA2_navigable da cui tutte dipendono: ovvero risultano definite come rdfs:subPropertyOf della corago:CPA2_navigable.

Individuata questa soluzione, il primo aspetto rilevante è la necessità di poter specificare quali tipi di entità ciascuna proprietà può mettere in relazione. Questo può essere fatto specificando le classi che rientrano rispettivamente nel dominio e nel codominio della relazione. Ovvero, quali saranno i tipi di entità che potranno comparire, rispettivamente, come soggetto e come oggetto della tripla avente come predicato la proprietà. In questo modo è possibile specializzare la proprietà in funzione dello specifico significato che assume in relazione ai termini che associa, indipendentemente dall’equivalenza della *property chain* sottostante. Ad esempio, è possibile specificare proprietà di navigazione distinte per differenziare la rappresentazione del titolo di un’opera dal titolo dello spettacolo, entrambe basate sulla stessa *property chain*, ma distinte dall’aver domini di tipo frbroo:F1_Work la prima, di tipo frbroo:F31_performance l’altra.

La possibilità di avere un controllo fine dell’ambito di applicazione della singola proprietà di navigazione consente di arricchire la sua definizione con altri attributi che ne precisino il valore semantico, la rilevanza nella descrizione dell’entità cui è riferita ed il contesto funzionale in cui deve essere utilizzata. Il formalismo utilizzato per estendere il modello descrittivo delle proprietà di navigazione è dato dalle proprietà di annotazione di OWL (owl:AnnotationProperty). Queste vengono impiegate nel linguaggio OWL DL per aggiungere metadati utili per arricchire la semantica degli assiomi definiti in una ontologia. Ad esempio, le

proprietà `rdfs:label` e `rdfs:comment` sono due *annotation properties* predefinite in OWL ed utilizzate per riportare nome e descrizione in linguaggio naturale di classi, proprietà ed individui. Le specifiche del linguaggio prevedono la possibilità di estenderne l'uso nel rispetto di alcuni vincoli⁴⁷.

Nella costruzione dell'ontologia Corago SM sono state quindi utilizzate le *annotation properties* come strumento per specificare le caratteristiche delle proprietà di navigazione che verranno sfruttate nell'implementazione dei requisiti funzionali.

In particolare, sono state introdotte delle *annotation properties* per:

1. indicare se la proprietà esprima un attributo identitario della classe cui è riferita;
2. specificare la rilevanza che la proprietà deve avere tra attributi della classe di riferimento;
3. specificare l'area tematica in cui rientra la proprietà;
4. individuare la presenza e posizione nella *property chain* di un nodo qualificante la relazione.

Nel paragrafo 2.4 è stata evidenziata la necessità di specificare a livello concettuale quali siano gli attributi identitari di una specifica classe. Questo livello di annotazione non è previsto dalle ontologie di riferimento essendo specifico del dominio di applicazione. Deve essere quindi introdotto all'interno dell'ontologia di dominio che costituisce un livello sovraordinato ed in particolare, seguendo il paradigma architetturale proposto, attraverso le proprietà di navigazione che costituiscono la rappresentazione calata sul dominio del modello concettuale sottostante.

E' stata introdotta a questo scopo la `corago:CPAN5_is_feature` (anch'essa una `owl:AnnotationProperty`): una proprietà di tipo booleano che "indica se la proprietà rientra tra quelle che specificano l'identità della risorsa". Ad esempio, le proprietà di navigazione che specificano il titolo di un'opera, il genere, i suoi autori riportano l'annotazione `corago:CPAN5_is_feature='true'`. Per contro non sono invece proprietà annotate come identitarie la relazione coi libretti.

Nel momento in cui è necessario confrontare due istanze appartenenti alla stessa classe è quindi possibile fare riferimento alle sole proprietà marcate come '*features*' per quella classe. Due istanze che presentino gli stessi valori per tutte le proprietà '*feature*' non possono essere distinte, devono essere quindi considerate equivalenti⁴⁸.

Nel modello concettuale definito dalle due ontologie di riferimento, tutti gli attributi previsti per descrivere una specifica classe di entità hanno pari dignità. Non è previsto un meccanismo per indicare la rilevanza che un attributo riveste nella descrizione di un oggetto. Questa limitazione viene superata con l'introduzione della *annotation property* `corago:CPAN1_priority`, di tipo numerico, che specifica la "rilevanza dell'attributo nella descrizione della risorsa".

Questa annotazione serve a definire l'ordine in base al quale le diverse proprietà di navigazione di una classe debbano essere considerate nel confrontare due istanze. Il criterio è direttamente derivato dalla rilevanza che le viene assegnata dal modello descrittivo dell'entità. Ad esempio, il titolo di un libretto avrà priorità maggiore rispetto all'anno di edizione. Lo stesso criterio può essere utilizzato per determinare l'ordinamento con cui le proprietà vengono elencate, ovvero l'articolazione della vista utente che rappresenta l'entità è

⁴⁷ Per il dettaglio relativo si rimanda a (W3C OWL Working Group, 2004) paragrafo 7.1.

⁴⁸ Si noti che in questo modo diventa possibile esprimere formalmente la completa equivalenza tra due risorse RDF all'interno del dominio di riferimento. In questo senso, due istanze che presentino gli stessi valori per tutte le proprietà identitarie definite dall'ontologia di dominio possono essere associate dal predicato `owl:sameAs` che indica la perfetta equivalenza di due risorse. Per un approfondimento sul tema si veda (Halpin et al., 2010).

organizzata a partire dalla priorità (o rilevanza) che è stata assegnata all'attributo attraverso la proprietà di navigazione che lo rappresenta.

A livello concettuale gli attributi descrittivi possono essere raggruppati in categorie distinte in funzione del loro contenuto informativo. Ad esempio, gli attributi identitari che riportano le informazioni essenziali per riconoscere un oggetto possono essere classificati come 'informazioni di base'. L'insieme delle informazioni relative alle rappresentazioni di un'opera sono invece riferibili alla sua 'cronologia'.

Questa organizzazione in aree tematiche degli attributi descrittivi è stata implementata attraverso l'annotazione `corago:CPAN4_thematic_area` che "definisce l'area tematica delle proprietà di navigazione".

I valori previsti per l'annotazione sono: "informazioni di base" (valore='basic_info'), "rappresentazioni" (valore='performance'), "personaggi" (valore='characters'), "contenuti" (valore='contents'), "struttura" (valore='structure'), "fonti" (valore='sources'), "riferimenti bibliografici" (valore='bibliography') e "relazioni" (valore='relations').

Il raggruppamento delle proprietà di navigazione in aree tematiche permette di selezionare i contenuti relativi ad una risorsa più aderenti al contesto d'uso. Ad esempio, in una anteprima sintetica è possibile limitare i contenuti alle sole proprietà relative alle 'informazioni di base' mentre, in un contesto più ampio dove assume rilevanza la presentazione dei documenti attinenti un'opera, la vista utente dovrebbe selezionare le aree "fonti", "riferimenti bibliografici" e "relazioni".

Si noti che le tre annotazioni fin qui introdotte: `corago:CPAN5_is_feature`, `corago:CPAN1_priority` e `corago:CPAN4_thematic_area`, permettono di organizzare le proprietà di navigazione in una rappresentazione dei contenuti costruita secondo le esigenze specifiche del dominio senza alterare minimamente il modello concettuale sottostante.

In alcuni casi particolari si rende necessario utilizzare contenuti informativi associati ed entità che compaiono in posizioni intermedie delle *property chain*. Sono i casi in cui una classe intermedia è portatrice di una informazione che qualifica l'intera relazione tra soggetto e oggetto del predicato. È il caso della relazione di responsabilità autoriale che viene qualificata attraverso la tipizzazione della classe intermedia `corago:C2_Actor_Role`. Diventa necessario, per poter segnalare questa rilevanza, specificare la classe che assume questo ruolo e la sua posizione all'interno della *property chain*.

A questo scopo sono state introdotte le due annotazioni: `corago:CPAN2_typing_class` e `corago:CPAN3_typing_class_position`. La prima indica la classe che specifica il tipo di relazione e la seconda la sua posizione assoluta. La funzione strumentale di questo tipo di annotazione è consentire il *parsing* della *property chain* per qualificare la proprietà di navigazione.

2.5.4 Costruzione delle viste utente

L'impianto fin qui descritto prevede che la rappresentazione dei contenuti, ovvero le viste per l'estrazione dei dati (*view*) presentati all'utente, siano costruite dal codice applicativo preposto (*controller*) a partire dalla semantica del modello concettuale (*model*). Di conseguenza il codice applicativo dovrà interpretare le componenti del modello concettuale in cui viene descritta la struttura e la rilevanza dei contenuti in termini funzionali. Questa parte, introdotta con lo strato delle proprietà di navigazione ed i relativi livelli di annotazione, viene utilizzata dal *controller* come strumento guida per organizzare i contenuti estratti dal *dataset* in viste utente.

In questo modo, l'applicazione di consultazione risulta svincolata dalla conoscenza pregressa della morfologia dei dati per permetterne la consultazione.

Nel complesso il *pattern* seguito ricalca l'impostazione tradizionale Model-View-Controller (MVC), ma cerca di ampliare il disaccoppiamento tra il *model* e il *controller*, ovvero profila per il *controller* un semplice ruolo di *adapter* tra il modello concettuale, la morfologia del dato e le viste che vengono presentate all'utente⁴⁹.

Nell'approccio proposto, è la semantica di modello che deve esporre al *controller/adapter* gli elementi necessari ad organizzare l'informazione; in questo senso parliamo di 'navigazione guidata dalla semantica di dominio'. Quindi, variazioni nell'organizzazione del modello concettuale, ad esempio riorganizzazione del *layer* intermedio delle proprietà navigabili, producono un effetto diretto sulla rappresentazione del dato. Ovvero è sufficiente modificare l'organizzazione della conoscenza descritta attraverso l'ontologia di dominio Corago SM per modificare l'organizzazione delle informazioni presentate all'utente finale. La conoscenza pregressa richiesta per l'implementazione dei componenti software che utilizzano il dataset è circoscritta alla capacità di interpretare le annotazioni dello strato intermedio e basarsi sulle *property chain* per accedere ai dati.

La procedura per la costruzione di una rappresentazione utente della risorsa <URI:01> è implementato attraverso il seguente algoritmo (espresso in pseudocodice) che vede l'interazione tra *controller/adapter* e *view*:

Passo 0: ricezione richiesta di generazione della *view* per la risorsa <URI:01>

Passo 1: identificazione della classe <CLASS:A> di appartenenza della risorsa <URI:01>

Passo 2: estrazione dall'ontologia di dominio dell'insieme [NP:CLASS:A] delle proprietà di navigazione (corago:CPA2_navigable) aventi per domain la classe <CLASS:A>

Passo 3: iterazione sull'insieme [NP:CLASS:A]

Passo 4: costruzione della query SPARQL in base alla *property path* dichiarata per la proprietà [NP:CLASS:A] [n-esima]

Passo 5: esecuzione della query SPARQL ed estrazione dei valori corrispondenti alla [NP:CLASS:A] [n-esima]

Passo 6: interpretazione delle annotazioni della proprietà [NP:CLASS:A] [n-esima]

Passo 7: applicazione dei criteri di rappresentazione applicabili ai valori estratti derivati dal punto 6

Passo 8: rappresentazione dei valori per la proprietà [NP:CLASS:A] [n-esima]

Passo 9: iterazione fino ad esaurimento elementi in [NP:CLASS:A]

Segue una illustrazione dei passaggi chiave dell'algoritmo:

Passi 1 e 2: i contenuti della vista prodotta sono in funzione della tipologia di risorsa consultata. È possibile quindi costruire delle viste specializzate per ciascuna entità del dominio.

⁴⁹ In questo senso il *pattern* adottato può essere classificato come un Model-View-Adapter (Zamudio Lopez et al., 2012).

Passi 4 e 5: attraverso l'equivalenza semantica tra proprietà di navigazione e *property chain* si disaccoppia il piano della rappresentazione dalla struttura del grafo che deve essere rappresentato.

Passi 6, 7 e 8: la rilevanza semantica della proprietà viene utilizzata per costruire la modalità di rappresentazione del dato estratto.

In funzione delle scelte di design architetturale dell'applicazione di consultazione, la suddivisione dei passi 6 e 7 tra *controller/adapter* e *view* può variare. Nell'implementazione del sistema sperimentale è stato seguito un approccio misto, ovvero il *controller* produce un oggetto già strutturato per la rappresentazione da parte della *view*, che però riporta anche tutta la struttura di annotazione delle proprietà. È quindi possibile inserire nella *view* criteri di rappresentazione basati sulla semantica del contenuto. Ad esempio, è possibile differenziare la rappresentazione in funzione dell'area tematica (derivata dall'annotazione

Esempio di *widget* per i personaggi di un'opera

CODICE HTML nella pagina JSP relativa alla scheda

```
<div style="margin-left:250px; white-space: nowrap; ">
<!-- Gestione della visualizzazione della proprietà: CPC4_has_character -->
<c:if test="${(aprop.getPropIri())=='http://corago.unibo.it/sm/CPC4_has_character'}">
<div id="characters1"></div>
<script>loadCPC4_has_character('#characters1','${results.getMainIri()}');</script>
</c:if>
</div>
```

Listato 2.2.1a

corago:CPAN4_thematic_area). In alternativa si possono creare dei componenti specifici per la rappresentazione di proprietà qualificate da una corago:CPAN2_typing_class: come la relazione tra opera e autore in cui è determinante evidenziare il tipo di responsabilità del singolo (per *Madama Butterfly*: compositore: Puccini, autori del libretto: Illica e Giacosa). I listati 2.1a e 2.2b riportano rispettivamente la *widget* della pagina relativa alla scheda generica delle entità e la funzione JavaScript per il caricamento asincrono e per la formattazione dei contenuti della proprietà di navigazione relativa ai personaggi di un'opera (corago: CPC4_has_character).

FUNZIONE JAVASCRIPT per il caricamento e formattazione dei contenuti

```
function loadCPC4_has_character(aDiv, aIRI) {
    var invCont = $(aDiv);
    var persNameList = "";
    var intTable = "<table width='90%'><thead><tr>
    <th>N.</th>
    <th style='text-align:left;'>Personaggio</th>
    <th style='text-align:left;'>Qualifica personaggio</th>
    <!--<th>Ruolo Vocale</th-->
    </tr></thead><tbody>";
    $.ajax({url : "${conf.getPublicUrlPrefix()}linkedTargetResource?IRI="
        + aIRI
        + "&property=http://corago.unibo.it/sm/CPC4_has_character",
        type : "GET",
        dataType : "xml",
        success : function(result) {
            data = $(result);
            if (data.find("resource").length > 0) {
                data.find("resource").each(
                    function() {
                        var targUrl = $(this).attr("propertyurl");
                        var pers = $(this).attr("propertyvalue");
                        persNameList += "%20" + pers;
                        var qual = $(this)
                            .find("tProp[propertyuri='http://corago.unibo.it/sm/CPC31_has_note']")
                            .text();
                        var rolVoc = getStringFromNodeList($(this)
                            .find("tProp[propertyuri='http://corago.unibo.it/sm/CPC16_character_type']"));
                        var num = $(this)
                            .find("tProp[propertyuri='http://corago.unibo.it/sm/CPC42_character_priority']")
                            .text();
                        intTable += ("<tr><td style='text-align: center;'>"
                            + num
                            + "</td><td><a href='"+targUrl+"'>"
                            + pers
                            + "</a></td><td>"
                            + qual
                            + "</td><!--<td>"
                            + rolVoc + "</td--><tr>");
                    }
                );
            }
            intTable += ("</tbody></table>");
            intTable += "<br><a href='${conf.getPublicUrlPrefix()}analyticalSearch?p="
                + persNameList
                + "' target='_blank'><b>ricerca opere correlate</b></a>";
            invCont.append(intTable);
            invCont.show();
        });
}
```

Listato 2.2.2b

La soluzione individuata si colloca nel filone di sviluppo di *framework* applicativi per la pubblicazione di contenuti RDF. La sperimentazione delle tecnologie semantiche per questo scopo si è sviluppata nel corso degli anni Duemila (Lausen et al., 2005) con l'obiettivo di realizzare applicazioni Web stand-alone per la navigazione di *dataset* modellati in base ad un set di ontologie di dominio. I servizi offerti sono: la

presentazione di schede descrittive delle entità principali del dominio trattato; la consultazione di liste ed indici ordinati e filtrabili in base a criteri prestabiliti; la ricerca full-text (Ding et al., 2010).

In alcuni ambiti applicativi, sono disponibili *framework* open source per la gestione dell'intero ciclo di vita del *dataset*, dalla fase di raccolta delle informazioni alla pubblicazione agli utenti finali.

Un esempio di questi è il *framework* Vitro⁵⁰ su cui è basata la soluzione VIVO per la pubblicazione dei risultati della ricerca svolta in ambito accademico (Börner et al., 2012). L'architettura Vitro/VIVO, essendo un progetto open source con una significativa diffusione e curato da una ampia comunità di sviluppo⁵¹, può essere un utile termine di paragone in relazione alle soluzioni adottate per la rappresentazione dei contenuti in formato HTML.

Entrambe le architetture prevedono la costruzione delle viste utente a partire da dati RDF estratti da un *endpoint* SPARQL e gestiti attraverso librerie software Java⁵². Le soluzioni divergono nella modalità di costruzione delle viste. In Vitro è stato adottato un meccanismo di prototipizzazione delle viste basate su un sistema di *templating*⁵³ che permette di specificare la struttura della vista associata ad una specifica proprietà di una classe. Al *template* HTML che definisce il *layout* che viene utilizzato per la formattazione della pagina HTML viene associata una interrogazione SPARQL per l'estrazione dei contenuti che verranno presentati. *Layout* e interrogazioni SPARQL vengono definite mediante file di configurazione che costituiscono il controllo utente (la *widget*) associato ad una specifica *data property*. Solo quest'ultima associazione viene riportata a livello di modello, ovvero all'interno dell'area del grafo utilizzata per la gestione delle configurazioni. La soluzione adottata da Vitro/VIVO confina la modalità di rappresentazione dei contenuti della base di conoscenza all'interno dell'applicazione di consultazione. La scelta di implementare la relazione tra proprietà della risorsa e sua rappresentazione in strati esterni al *dataset* fa sì che il modello in base al quale si è definita la rappresentazione dei contenuti non sia né accessibile né intelligibile da terzi.

Al contrario, la soluzione proposta e sperimentata in questo studio delega allo strato applicativo di presentazione solo il ruolo di gestione del *layout* della pagina HTML. L'accesso ai contenuti è interamente mediato dallo strato intermedio delle proprietà di navigazione, che, contrariamente alla soluzione Vitro/VIVO fa parte a pieno titolo della base di conoscenza.

D'altra parte, questa soluzione non esclude un certo grado di accoppiamento tra vista e modello. Come illustrato nel listato 2.1a/b, la costruzione di rappresentazioni HTML di specifiche proprietà richiede che il *controller* fornisca i contenuti alla vista della proprietà facendovi riferimento in modo esplicito. Ovvero, se si vuole ottenere un *layout* particolare per mostrare, ad esempio, l'elenco dei personaggi dell'opera, in fase di costruzione della pagina HTML, il *controller* dovrà delegare la rappresentazione ad una *view* specializzata sulla proprietà di navigazione corrispondente.

Allo stato attuale, questa forma di 'delega' è inserita all'interno delle *view*, implementate come Java Server Pages (JSP), utilizzando lo *scripting* JavaScript. Se per una proprietà è prevista una rappresentazione specializzata, viene invocata la funzione dello script corrispondente che si occupa di costruire il *layout* HTML.

⁵⁰ Il codice sorgente della soluzione è pubblicato su GitHub all'indirizzo: <https://github.com/vivo-project/Vitro> (consultato 30/06/2019).

⁵¹ Il progetto open-source VIVO è mantenuto dalla comunità Duraspace: <https://duraspace.org/vivo> (consultato 30/06/2019).

⁵² Nello specifico sia Corago LOD che il *framework* Vitro/VIVO impiegano la libreria Apache JENA.

⁵³ Nello specifico è stato adottato il *framework* di gestione di templating HTML Apache Freemarker.

Qualora invece si volesse introdurre un meccanismo di *templating* come Apache Freemarker, il *controller* dovrebbe accedere ad una configurazione che associa la proprietà al template specializzato per la sua visualizzazione. Tale configurazione non dovrebbe comunque essere riportata a livello di modello in quanto strettamente legato alla presentazione del contenuto nel contesto applicativo del sistema di consultazione. Per quanto possibile va mantenuta la separazione tra elementi relativi al valore semantico dell'attributo e le modalità di rappresentazione derivate da essi. I primi infatti devono rientrare nel perimetro del modello concettuale, le seconde restano comunque delegate alle applicazioni di consultazione.

2.6 Indicizzazione dei contenuti testuali

L'archivio *Corago* è ricco di contenuti testuali che possono essere efficacemente utilizzati come chiavi di accesso al patrimonio informativo a patto che vengano organizzati e strutturati sotto forma di riferimenti alle risorse che descrivono. Nel paragrafo 1.5.2 è stata prevista la possibilità per l'utente di effettuare ricerche testuali senza dover necessariamente specificare il tipo di entità ed il relativo attributo a cui i criteri di ricerca debbano essere applicati.

Per raggiungere questo risultato è necessario che ciascuna risorsa venga indicizzata in base ad un preciso insieme di contenuti derivati dai suoi attributi testuali. Questo insieme di testi deve conservare un riferimento alla natura degli attributi da cui viene estratto. Il testo così ottenuto costituisce un 'documento' associato alla risorsa che la 'descrive' e può essere utilizzato dal motore di ricerca per individuarla in base a criteri dati.

In fase di presentazione dei risultati, l'indice deve consentire di associare le porzioni di testo presenti nel 'documento' che corrispondono ai termini utilizzati come criteri per la ricerca. L'eventuale riferimento alla natura dell'attributo da cui il testo deriva serve a chiarire la relazione con la risorsa individuata e quindi la sua rilevanza come risultato della ricerca. Nel caso di risorse in formato RDF, la semantica della relazione tra attributo testuale e la risorsa è definita in modo formale a priori. Materialmente è rappresentata dal grafo che collega la risorsa al testo confluito nell'indice.

La costruzione di un siffatto indice testuale, nel caso di risorse RDF, può sfruttare la possibilità di selezionare i contenuti testuali in base alla semantica delle loro relazioni. In modo simmetrico, la stessa potrà essere utilizzata in fase di presentazione dei risultati per specificare la natura della relazione tra testo e risorsa⁵⁴.

In base al requisito [REQ4] il sistema deve essere in grado di ricercare una risorsa a partire dal contenuto testuale di alcune sue proprietà individuate in modo esplicito sia in fase di costruzione dell'indice (una selezione dei suoi contenuti), sia nella presentazione dei risultati all'utente. Il vantaggio immediato per l'utente consiste nel poter valutare direttamente la rilevanza degli attributi in base al quale il singolo risultato è stato selezionato. Inoltre, essendo navigabili le relazioni tra il risultato e le risorse che rappresentano gli attributi confluiti nell'indice, è sempre possibile accedere alla relativa scheda per approfondirne il contenuto.

L'ordinamento dei risultati, ovvero il *ranking*, potrà essere basato in prima istanza sulla rilevanza assoluta dei termini utilizzati per la ricerca e, in prospettiva, in base ad un peso calcolato in funzione della rilevanza della relazione che lega il testo individuato al risultato. Ad esempio, il termine 'Roma' assume maggior rilevanza come località di uno spettacolo che come luogo di edizione di un documento e addirittura nulla rispetto al luogo di nascita di un interprete. In base a questo criterio lo spettacolo dovrà precedere nell'ordinamento il libretto.

2.6.1 Articolazione del requisito funzionale

Per raggiungere gli obiettivi previsti dal [REQ4] il sistema dovrà avere le seguenti caratteristiche:

1. essere in grado di indicizzare le tre entità di base, opere, eventi e documenti attraverso i relativi contenuti testuali;
2. permettere di specificare gli attributi che devono essere inclusi nel perimetro dell'indice e riportarne la natura della relazione;

⁵⁴ Per una panoramica sugli strumenti e le metodologie di indicizzazione dei contenuti in formato RDF si rinvia a (Luo et al., 2012).

3. evidenziare il contenuto testuale corrispondente ai criteri di ricerca specificati per ciascun risultato;
4. restituire il ‘contesto informativo’ di riferimento del singolo risultato dandone giustificazione;
5. permettere di valutare l’efficacia del criterio e la rilevanza del risultato navigando direttamente le risorse riportate nel contesto.

Sulla base di questi requisiti è stata condotta l’analisi tecnica e l’implementazione dell’indice descritta nel paragrafo seguente.

2.6.2 Analisi tecnica ed implementazione dell’indice

2.6.2.1 Stato dell’arte e strumenti disponibili

La libreria software Apache Lucene può essere considerata come uno strumento di riferimento per l’indicizzazione e al ricerca di contenuti testuali (Smiley et al., 2015, p. 2). La libreria consente di indicizzare i testi presenti in strutture chiave-valore (*documents*) che vengono sottoposte a tokenizzazione e *stemming* per produrre un indice di termini (*terms*). L’indice risultante, un indice inverso, contiene per ciascun termine il riferimento alla posizione dei *documents* in cui compare.

La libreria consente ricerche in base all’indice dei termini estratti da tutti i documenti in cui questi compaiono assegnando a ciascuno di essi una rilevanza. La rilevanza di un documento in relazione ai termini ricercati è di norma calcolata a partire dalla *term frequency–inverse document frequency* (TF-IDF). Ottengono quindi più rilevanza quei documenti in cui i termini ricercati compaiono con più frequenza rispetto alla loro diffusione su tutto il corpus (Martin e Jurafsky, 2009, p. 805).

La libreria Lucene viene impiegata da diversi motori di indicizzazione come Apache SOLR⁵⁵ o Elasticsearch⁵⁶. La stessa è utilizzata in Blazegraph, il *repository* di triple impiegato nel sistema sperimentale, per operare ricerche full-text all’interno di interrogazioni SPARQL. Questa funzionalità viene implementata attraverso una serie di predicati definiti sotto un apposito namespace: <http://www.bigdata.com/rdf/search#>. In particolare il predicato `bds:search` permette di individuare tutte le occorrenze di nodi di tipo *literal* inseriti nell’indice che corrispondono al criterio testuale specificato.

Ad esempio il predicato: `?label bds:search "Otello"` individuerà tutti i nodi di tipo *literal* che contengono il termine proiettandoli sulla variabile `?label`. Combinandolo col predicato `?s rdfs:label ?label`, si individueranno tutte le risorse (`?s`) che hanno una etichetta (definita come proprietà di tipo `rdfs:label`) in cui compare il termine “Otello”.

Il testo che viene inserito nell’indice è derivato da tutti i nodi RDF di tipo *literal* presenti nella posizione di oggetto di tutte le triple caricate nel *repository*. Questa soluzione permette di ricercare qualsiasi contenuto testuale associato ad ogni risorsa, ovvero nodo in posizione di soggetto di una tripla. Individuata la risorsa che corrisponde ad un criterio testuale è possibile navigare il suo grafo per risalire all’entità oggetto della ricerca. Ad esempio se sono interessato ad individuare tutte le opere ambientate in una “amena campagna” dovrò identificare le ambientazioni (entità di tipo: `corago:C10_Scene_Setting`) nella cui descrizione compare il criterio selezionato e risalire alle relative opere (entità di tipo: `frbroo:F1_Work`) attraverso la proprietà `corago:CPC65_work_scene_setting`.

⁵⁵ Per approfondimenti: <http://lucene.apache.org/solr/> (consultato 30/08/2019).

⁵⁶ Per approfondimenti: <https://www.elastic.co/products/elasticsearch> (consultato 30/08/2019).

L'implementazione del requisito [REQ4] richiede inoltre di circoscrivere il perimetro della ricerca ad un sottoinsieme dei contenuti testuali riferiti ad una risorsa e di esplicitarne la relazione con la stessa sia in fase di ricerca che in fase di visualizzazione del risultato. Questo comporta la necessità di definire a priori la struttura dei grafi che associano il nodo di tipo *literal* contenente il testo e risorsa che dovrà essere individuata come risultato. Ovvero, solo il sottoinsieme di nodi con contenuto testuale specificato dai grafi selezionati dovrà essere oggetto della ricerca. Di conseguenza, il grafo che collega il risultato al testo indicizzato deve essere applicato come clausola di restrizione nelle query SPARQL che utilizzano l'indice full-text. Questo aumenta in modo significativo la complessità della clausola di restrizione della query in modo tale da rendere il piano di esecuzione dipendente dalla distanza tra risorsa e nodo indicizzato.

Una possibile alternativa è selezionare a priori i contenuti testuali in base ai quali indicizzare una risorsa⁵⁷. Questa soluzione è adottata ad esempio dal già citato *framework* Vitro⁵⁸. In questo caso l'architettura di sistema prevede comunque l'introduzione di un server Apache SOLR come motore di indicizzazione separato dal *repository* RDF. L'approccio seguito è di utilizzare delle query SPARQL in fase di indicizzazione per estrarre i contenuti specifici di ciascuna tipologia di risorsa da inserire nei *document*. La ricerca produrrà dei risultati in base al contenuto testuale confluito in essi e l'applicazione di consultazione presenterà l'elenco dei risultati rinviando alla scheda della singola risorsa per la consultazione dei contenuti.

In questa architettura, l'impiego di un motore esterno comporta la perdita della semantica della relazione tra la risorsa indicizzata ed il testo ricercato. L'estrazione mediante query SPARQL dei testi e la loro trasposizione in strutture di tipo chiave-valore impedisce infatti di conservare il nesso tra occorrenza testuale e risorsa individuata. Viene persa inoltre l'identità (URI) del nodo cui il testo è direttamente associato essendo il testo direttamente riferito all'entità che si vuole produrre come risultato. Ad esempio, viene perso il riferimento a Venezia come *frbroo: F9_Place* quando si cercano gli spettacoli andati in scena su quella piazza.

Seguendo questo approccio, verrebbe soddisfatto il requisito di poter selezionare gli attributi che contribuiscono alla costruzione dell'indice, ma non di riuscire a specificare la natura della relazione, la sua rilevanza e l'entità oggetto della relazione.

2.6.2.2 Selezione dei contenuti e dei risultati

La selezione dei contenuti testuali che dovranno confluire nell'indice deve partire dalla rilevanza che questi assumono nella descrizione delle diverse tipologie di entità. Ad esempio, come già osservato, il luogo di prima rappresentazione di un'opera è più significativo del luogo di nascita del suo autore. Entrambe le relazioni hanno come estremi le stesse tipologie di entità, ma cambiano semantica e morfologia del grafo che rappresenta la relazione. Per selezionare i contenuti testuali da indicizzare è quindi necessario specificare la relazione, espressa in forma di grafo, che intercorre tra la classe del risultato e il nodo di tipo *literal* che contiene il testo.

Per consentire l'individuazione di quali proprietà indicizzare è stata introdotta una classe di proprietà tra oggetti (ovvero definite come OWL *object properties*) all'interno dell'ontologia di dominio. La loro funzione è specificare la relazione tra la classe del risultato, specificata come dominio, e il contenuto testuale da indicizzare, indicato come *range*, espressa sotto forma di *property path*.

⁵⁷ Un utile termine di paragone alla soluzione qui presentata è l'approccio al problema proposto da (Bast e Buchhold, 2013).

⁵⁸ Vedi paragrafo 2.5.4.

Le proprietà definite a questo scopo vengono fatte derivare da una proprietà astratta denominata `corago:CPA4_ft_breadcrumb`⁵⁹. Questa dipendenza ne indica la finalità: individuare i contenuti testuali che dovranno permettere l'individuazione dei risultati per le diverse tipologie di entità. Come risultato si avrà che solo le entità indicizzate in base a queste potranno comparire nei risultati delle ricerche testuali.

Esempio di definizione per la proprietà `corago:CPBC21_published_by`

```
<!-- http://corago.unibo.it/sm/CPBC21_published_by -->

<owl:ObjectProperty rdf:about="http://corago.unibo.it/sm/CPBC21_published_by">
  <rdfs:subPropertyOf rdf:resource="http://corago.unibo.it/sm/CPA4_ft_breadcrumb"/>
  <rdfs:domain rdf:resource="http://erlangen-
crm.org/efrbroo/F3_Manifestation_Product_Type"/>
  <rdfs:range rdf:resource="http://erlangen-crm.org/current/E39_Actor"/>
  <owl:propertyChainAxiom rdf:parseType="Collection">
    <rdf:Description rdf:about="http://erlangen-crm.org/efrbroo/CLR6_should_carry"/>
    <rdf:Description>
      <owl:inverseOf rdf:resource="http://erlangen-crm.org/current/P94_has_created"/>
    </rdf:Description>
    <rdf:Description rdf:about="http://erlangen-crm.org/current/P14_carried_out_by"/>
    <rdf:Description rdf:about="http://erlangen-crm.org/current/P131_is_identified_by"/>
  </owl:propertyChainAxiom>
  <owl:propertyChainAxiom rdf:parseType="Collection">
    <rdf:Description>
      <owl:inverseOf rdf:resource="http://erlangen-
crm.org/efrbroo/R26_produced_things_of_type"/>
    </rdf:Description>
    <rdf:Description>
      <owl:inverseOf rdf:resource="http://erlangen-crm.org/current/P94_has_created"/>
    </rdf:Description>
    <rdf:Description rdf:about="http://erlangen-crm.org/current/P14_carried_out_by"/>
    <rdf:Description rdf:about="http://erlangen-crm.org/current/P131_is_identified_by"/>
  </owl:propertyChainAxiom>
  <rdfs:comment xml:lang="it">Editore</rdfs:comment>
  <rdfs:comment xml:lang="en">Publisher</rdfs:comment>
</owl:ObjectProperty>
```

Listato 2.3

Il controllo delle tipologie di entità ammesse come risultato è ottenibile in due modi: in modo esplicito specificando il dominio cui vengono applicate le proprietà *breadcrumbs*, oppure in modo implicito. Infatti, dovendo i risultati necessariamente corrispondere ai tipi che compaiono in posizione di soggetto nei grafi

⁵⁹ Il nome scelto per la proprietà fa riferimento alla fiaba di Pollicino. Come le briciole di pane (*breadcrumbs* in inglese), così queste proprietà aiutano l'utente a ripercorrere la strada che porta dal testo ricercato al risultato individuato dal motore di ricerca.

identificati dalle *property path*, le tipologie per cui non è previsto almeno un *breadcrumb* non potranno comparire tra i risultati. La rilevanza di questa possibilità dal punto di vista dell'utente verrà approfondita nel paragrafo 3.1.

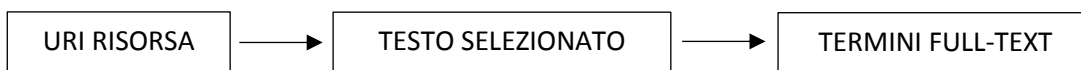
2.6.2.3 Morfologia dell'indice

Una volta identificati i contenuti testuali che devono essere inclusi nell'indice questi devono essere strutturati in modo da poter essere indicizzati dal *repository*. Come descritto nel paragrafo precedente, il *repository* Blazegraph prevede l'indicizzazione dei nodi RDF di tipo *literal* in posizione di oggetto di tutte le triple del *dataset*. L'insieme dei contenuti selezionati per l'indicizzazione per ciascuna risorsa in base alle proprietà *breadcrumbs* devono essere raccolti in un nodo *literal* associato direttamente al nodo che rappresenta la risorsa indicizzata. La proprietà specifica `corago:CPA5_ft_indexed_by` associa i due nodi.

La struttura delle triple utilizzate per l'indicizzazione è la seguente:

URI risorsa – `corago:CPA5_ft_indexed_by` – testo selezionato

Pertanto, l'indicizzazione da parte del *repository* viene applicata al nodo *literal* in posizione di oggetto nella tripla e genera l'indice dei termini rilevanti estratti dal contenuto testuale della stringa.



Una volta generate ed indicizzate le triple così strutturate, diventa possibile effettuare ricerche sullo specifico sottoinsieme di contenuti testuali selezionati a questo scopo. La ricerca può essere effettuata con una query SPARQL che utilizzi lo schema di grafo seguente:

Template di grafo per l'utilizzo dell'indice full-text (pseudocodice SPARQL)

```
...  
  
?s corago:CPA5_ft_indexed_by ?text.  
  
?text bds:search "Otello"  
  
...
```

Listato 2.4

Nell'esempio riportato nel listato 2.4, la variabile `?s` rappresenta la risorsa indicizzata su cui verranno proiettate le occorrenze selezionate dalla query; la variabile `?text` è l'oggetto della ricerca attivata dal predicato speciale per la ricerca full-text della stringa "Otello", ovvero il criterio di ricerca specificato dall'utente. Il risultato della query sarà costituito da tutte le risorse nel cui indice testuale compare il termine ricercato.

La soluzione proposta risponde in modo sostenibile al requisito di poter effettuare ricerche delle risorse dell'archivio in base a specifici contenuti testuali. Non è però sufficiente a restituire il contesto di provenienza del testo in base al quale ciascun risultato è stato individuato.

Nella composizione del nodo *literal* vengono infatti presi in considerazione solo i contenuti testuali derivati dalle proprietà '*breadcrumbs*'. In questo modo viene omesso il riferimento alla risorsa oggetto della

relazione: nell'esempio riportato sopra, viene considerata la stringa "Venezia" ma si perde il fatto che si tratta del toponimo relativo alla città di Venezia (risorsa definita come `frbroo:F9_Place`). Viene prelevato il contenuto, ma se ne perde il 'contesto di origine'. Questo ha come ulteriore conseguenza che sarà impossibile ricostruire la relazione che intercorre tra il risultato e il termine "Venezia" che ha permesso di individuarlo come tale. Ovvero, in questo esempio, si perde il riferimento al grafo che collega l'opera al luogo di prima rappresentazione, grafo identificato dalla relativa proprietà '*breadcrumb*'.

È quindi necessario mantenere il riferimento alla risorsa da cui il testo è stato derivato. Per questo è sufficiente prevedere l'inserimento oltre al contenuto del nodo *literal* anche degli URI relativi alle risorse sotto forme di coppie chiave-valore. La struttura risultante può essere rappresentata come una concatenazione di URI e stringhe separata da opportuni caratteri convenzionali. Questa struttura può essere serializzata all'interno del nodo *literal* e quindi indicizzata. In fase di estrazione dei risultati, il contenuto del nodo verrà sottoposto a *parsing* per restituire l'insieme delle coppie chiave-valore contenute.

Con questo accorgimento è possibile ricostruire a posteriori la natura della relazione tra il testo ed il risultato andando a ricercare la proprietà '*breadcrumb*' che unisce l'URI del risultato e l'URI della risorsa da cui il testo è stato derivato. In questo modo sarà sempre possibile ricostruire il contesto di origine delle corrispondenze individuate dalla ricerca full-text nei contenuti testuali riferiti ai singoli risultati prodotti.

Ad ulteriore integrazione dei contenuti delle triple dedicate alla ricerca testuale, in fase di sperimentazione si è dimostrato necessario aggiungere una serie di attributi del risultato dedicati alla gestione della navigazione dei risultati. Si tratta di attributi identitari⁶⁰ della risorsa indicizzata necessari per poter ordinare l'insieme dei risultati e che risulterebbe oneroso caricare singolarmente a valle dell'estrazione. Questi vengono inseriti in una sezione distinta del nodo *literal* che viene concatenata alla prima attraverso l'uso di una sequenza convenzionale di caratteri di separazione (ad es. `###`). Questa seconda sezione verrà utilizzata in fase di presentazione dei risultati per accelerare le funzionalità di paginazione ed ordinamento dei risultati⁶¹.

2.6.2.4 Costruzione dell'indice

La procedura di generazione delle triple che costituiscono l'indice testuale delle risorse è implementabile attraverso il seguente algoritmo (espresso in pseudocodice):

Passo 0: Individuazione di tutte le proprietà *breadcrumb* definite come sotto-proprietà della `corago:CPA4_ft_breadcrumb`.

Passo 1: Individuazione dei tipi definiti come dominio delle proprietà *breadcrumb* individuate al passo 0.

Passo 2: Creazione di un insieme temporaneo di risorse afferenti ai tipi individuati al passo 1.

Passo 3: Estrazione per ciascuna delle risorse nell'insieme creato al passo 2 dei contenuti testuali per in base alle proprietà *breadcrumb* aventi il tipo della risorsa come dominio.

Passo 4: Composizione della stringa che verrà indicizzata come concatenazione degli URI delle risorse e dei relativi testi (coppie chiave-valore).

Passo 5: Estrazione delle proprietà per l'ordinamento dei risultati.

⁶⁰ Fare riferimento al paragrafo 2.5 per una definizione delle diverse tipologie di attributo.

⁶¹ Si rinvia alla appendice C.1 per una trattazione degli aspetti implementativi adottati nella soluzione sperimentale.

Passo 6: Creazione della tripla che associa la stringa indicizzata al risultato con la proprietà specifica `corago:CPA5_ft_indexed_by`.

Passo 7: Inserimento nel *repository* mediante query SPARQL di UPDATE.

Passo 8: Iterazione sul passo 3 fino ad esaurimento dell'insieme di risorse estratto al passo 2.

Al termine della procedura, per ciascuna risorsa estratta al passo 2, sarà presente nel *repository* una tripla il cui nodo *literal* in posizione di oggetto viene indicizzato dal servizio di indicizzazione full-text del *repository*. Nel caso del *repository* adottato nel progetto, l'indicizzazione avviene contestualmente al caricamento della tripla. La modalità di indicizzazione e le sue caratteristiche, così come le funzionalità in fase di ricerca sono dipendenti dall'implementazione del servizio prevista da ciascun *repository*. Va osservato però che anche in assenza delle funzionalità di indicizzazione full-text, resta comunque possibile utilizzare le funzionalità di ricerca di testi basate sulle espressioni regolari che sono previste come parte delle raccomandazioni del W3C per l'implementazione di SPARQL 1.1. La soluzione proposta per l'implementazione del requisito [REQ4] resta applicabile a tutti i sistemi che implementino il protocollo.

2.6.2.5 Utilizzo dell'indice

La ricerca a testo libero, nel sistema così disegnato, si compone di due fasi: la prima di ricerca vera e propria dei risultati corrispondenti ai criteri inseriti, la seconda di ricostruzione del contesto dei riferimenti testuali individuati. Entrambe sono implementate attraverso query SPARQL di cui seguono gli esempi prototipali (listato 2.5).

Template di query SPARQL per la ricerca in base a criteri testuali

```
select ?s ?text {  
  
  ?s corago:CPA5_ft_indexed_by ?text.  
  
  ?text bds:search "Otello"}
```

Template di query SPARQL per l'estrazione del contesto dei riferimenti ai criteri

```
Select ?s ?p ?o {  
  
  <URI risultato> ?p <URI riferimento>.  
  
  ?p rdfs:subPropertyOf corago:CPA4_ft_breadcrumb}
```

Listato 2.5

L'esempio evidenzia come le query per la ricerca testuale utilizzino l'implementazione delle funzionalità di ricerca full-text del *repository* attraverso l'uso del predicato specifico *"bds:search"* che viene però applicato al contenuto del nodo dedicato all'indicizzazione delle risorse individuato dallo specifico predicato: `corago:CPA5_ft_indexed_by`. In questo modo la ricerca resta circoscritta all'interno dell'indice appositamente costruito. Ne consegue un pieno controllo sulle tipologie di risultati derivanti dalla ricerca: solo le risorse indicizzate in base alle proprietà utilizzate verranno incluse nei risultati. Al tempo stesso l'insieme delle risorse indicizzate è finito e la numerosità degli elementi presenti è nota a valle del processo di indicizzazione. Il costo della query principale per le ricerche testuali, così come delle query che impiegano

l'indice come criterio aggiuntivo, diventa prevedibile e costante⁶². È quindi possibile impiegare l'indice in unione ad altri criteri di ricerca come i vincoli rispetto alla coordinata temporale o criteri di restrizione geografica (si vedano i casi d'uso che sfruttano questa possibilità descritti al paragrafo 5.2).

⁶² Il costo è proporzionale infatti al numero di predicati necessari a definire il criterio di restrizione e al numero di elementi selezionati da essi ovvero, nel nostro caso, alle risorse inserite nell'indice.

Esempio di tripla dell'indice full-text

<<http://corago.unibo.it/resource/BRANI/0000350175>>

<http://corago.unibo.it/sm/CPA5_ft_indexed_by>

"[http://corago.unibo.it/resource/RESPONS/BPP008001600&&Barlocchi, Giovanni Gualberto](http://corago.unibo.it/resource/RESPONS/BPP008001600&&Barlocchi,%20Giovanni%20Gualberto)####[http://corago.unibo.it/resource/RESPONS/Z00004310400&&Latilla, Gaetano](http://corago.unibo.it/resource/RESPONS/Z00004310400&&Latilla,%20Gaetano)####[http://corago.unibo.it/resource/BRANI/AMBIENTAZIONE/0000350175&&La scena \xc3\xa8 in Firenze](http://corago.unibo.it/resource/BRANI/AMBIENTAZIONE/0000350175&&La%20scena%20%C3%A8%20in%20Firenze)####[http://corago.unibo.it/resource/BRANI/0001381476&&Madama Ciana](http://corago.unibo.it/resource/BRANI/0001381476&&Madama%20Ciana)####[http://corago.unibo.it/resource/BRANI/0001381548&&Madama Ciana](http://corago.unibo.it/resource/BRANI/0001381548&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/0000350186&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/0000350186&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/0000350192&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/0000350192&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/0000350198&&Gli *artigiani arricchiti](http://corago.unibo.it/resource/SPETTACOLI/0000350198&&Gli%20artigiani%20arricchiti)####[http://corago.unibo.it/resource/SPETTACOLI/0000379189&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/0000379189&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/0000389113&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/0000389113&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/SST0026499&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/SST0026499&&Madama%20Ciana)####<http://corago.unibo.it/resource/PERSONAG/0000801554&&Marzia>####<http://corago.unibo.it/resource/PERSONAG/0000801555&&Orazio>####<http://corago.unibo.it/resource/PERSONAG/0000801556&&Panicone>####[http://corago.unibo.it/resource/PERSONAG/0000801557&&madama Ciana](http://corago.unibo.it/resource/PERSONAG/0000801557&&madama%20Ciana)####<http://corago.unibo.it/resource/PERSONAG/0000801558&&Sfrappa>####<http://corago.unibo.it/resource/PERSONAG/0000801559&&Fiammetta>####<http://corago.unibo.it/resource/PERSONAG/0000801560&&Moschino>####<http://corago.unibo.it/resource/PERSONAG/0000801561&&Sgrana>####[http://corago.unibo.it/resource/SPETTACOLI/0000379552&&L' *ambizione delusa](http://corago.unibo.it/resource/SPETTACOLI/0000379552&&L%20ambizione%20delusa)####<http://corago.unibo.it/resource/SPETTACOLI/0000379569&&Ciana>####<http://corago.unibo.it/resource/SPETTACOLI/0000887958&&Ciana>####[http://corago.unibo.it/resource/SPETTACOLI/0001381525&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/0001381525&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/0000382264&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/0000382264&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/0000403097&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/0000403097&&Madama%20Ciana)####<http://corago.unibo.it/resource/SPETTACOLI/DATA/0000350192&&1738/02>####[http://corago.unibo.it/resource/SPETTACOLI/0000657196&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/0000657196&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/SST0010021&&La *Ciana](http://corago.unibo.it/resource/SPETTACOLI/SST0010021&&La%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/SST0026505&&Madama Ciana](http://corago.unibo.it/resource/SPETTACOLI/SST0026505&&Madama%20Ciana)####[http://corago.unibo.it/resource/SPETTACOLI/BRA/FORMA/0000350179&&dramma giocoso per musica](http://corago.unibo.it/resource/SPETTACOLI/BRA/FORMA/0000350179&&dramma%20giocosopermusica)####[http://corago.unibo.it/resource/BIBLIOGRAFIA/0000099574&&Grove Music Online](http://corago.unibo.it/resource/BIBLIOGRAFIA/0000099574&&Grove%20Music%20Online)####<http://corago.unibo.it/resource/RESPONS/LUOGO/MOR/Z00004310400&&Napoli>####<http://corago.unibo.it/resource/RESPONS/LUOGO/NAS/Z00004310400&&Bari>####[http://corago.unibo.it/resource/DOCUMENTI/0000477050&&Madama Ciana](http://corago.unibo.it/resource/DOCUMENTI/0000477050&&Madama%20Ciana)####[http://corago.unibo.it/resource/DOCUMENTI/DRT0002089&&L' *ambizione delusa](http://corago.unibo.it/resource/DOCUMENTI/DRT0002089&&L%20ambizione%20delusa)####[http://corago.unibo.it/resource/DOCUMENTI/DRT0005561&&Gli *artigiani arricchiti](http://corago.unibo.it/resource/DOCUMENTI/DRT0005561&&Gli%20artigiani%20arricchiti)####<http://corago.unibo.it/resource/DOCUMENTI/DRT0010017&&Ciana>####<http://corago.unibo.it/resource/DOCUMENTI/DRT0010019&&Ciana>####[http://corago.unibo.it/resource/DOCUMENTI/DRT0010021&&La *Ciana](http://corago.unibo.it/resource/DOCUMENTI/DRT0010021&&La%20Ciana)####[http://corago.unibo.it/resource/DOCUMENTI/DRT0026491&&Madama Ciana](http://corago.unibo.it/resource/DOCUMENTI/DRT0026491&&Madama%20Ciana)####<http://corago.unibo.it/resource/PERSONAG/QUAL/servo&&servo>####[http://corago.unibo.it/resource/PERSONAG/QUAL/loro_serva&&loro serva](http://corago.unibo.it/resource/PERSONAG/QUAL/loro_serva&&loro%20serva)####<http://corago.unibo.it/resource/PERSONAG/QUAL/adulatore&&adulatore>####[http://corago.unibo.it/resource/PERSONAG/QUAL/cavaliere_della_Luna_Calante&&cavaliere della Luna Calante](http://corago.unibo.it/resource/PERSONAG/QUAL/cavaliere_della_Luna_Calante&&cavaliere%20della%20Luna%20Calante)####[http://corago.unibo.it/resource/PERSONAG/QUAL/e_di_%5BSfrappa%5D&&e di \[Sfrappa\]](http://corago.unibo.it/resource/PERSONAG/QUAL/e_di_%5BSfrappa%5D&&e%20di%20Sfrappa)####[http://corago.unibo.it/resource/PERSONAG/QUAL/gentildonna_povera%2C_amante_di_Orazio&&gentil donna povera, amante di Orazio](http://corago.unibo.it/resource/PERSONAG/QUAL/gentildonna_povera%2C_amante_di_Orazio&&gentilidonna%20povera,%20amante%20di%20Orazio)####[http://corago.unibo.it/resource/PERSONAG/QUAL/gentiluomo_amante_di_Marzia&&gentiluomo amante di Marzia](http://corago.unibo.it/resource/PERSONAG/QUAL/gentiluomo_amante_di_Marzia&&gentiluomo%20amante%20di%20Marzia)####[http://corago.unibo.it/resource/PERSONAG/QUAL/padre_di_%5Bmadama_Ciana%5D&&padre di \[madama Ciana\]](http://corago.unibo.it/resource/PERSONAG/QUAL/padre_di_%5Bmadama_Ciana%5D&&padre%20di%20madama%20Ciana)####[http://corago.unibo.it/resource/LEGAMI/LLO/0000350175-0001381476&&1 tratto da](http://corago.unibo.it/resource/LEGAMI/LLO/0000350175-0001381476&&1%20tratto%20da)####<http://corago.unibo.it/sm/indexingAuthor&&Latilla> [...]

Listato 2.6

Le triple identificate in base al predicato `corago:CPA5_ft_indexed_by` presentano come soggetto la risorsa individuata e come risultato ed in posizione oggetto la stringa che deve essere utilizzata dall'applicazione di consultazione per estrarre gli attributi testuali indicizzati e quelli utili per la gestione dell'ordinamento. Queste triple contengono le informazioni di base necessarie per permettere alle viste utente di organizzare i risultati in elenchi paginabili ed ordinabili (il listato 2.6 ne riporta un esempio).

L'individuazione delle porzioni di testo corrispondenti ai criteri di ricerca (corrispondenze) avviene confrontando i valori delle coppie chiave-valore di ciascun risultato con i criteri immessi. Sulla base di un confronto della superficie lessicale, che riproduce il criterio impiegato effettivamente dal motore di indicizzazione, è possibile individuare le porzioni di testo rilevanti ed il corrispondente URI della risorsa da cui derivano. Vengono così ricostruite le corrispondenze tra risultato e quegli attributi in base ai quali è stato individuato. A partire da questo elenco di corrispondenze, in funzione dei requisiti di consultazione, viene infine recuperata anche la natura della relazione tra risultato e singolo attributo indicizzato.

L'applicazione di consultazione è quindi in grado di restituire nell'ordine:

1. il riferimento al risultato e i suoi attributi descrittivi di base;
2. gli attributi testuali con le rispettive corrispondenze individuate;
3. la natura delle relazioni tra questi ed il risultato.

L'insieme di questi tre livelli informativi costituisce quello che definiamo il 'contesto informativo' del risultato, ovvero l'insieme delle informazioni necessarie all'utente per valutare la sua rilevanza in funzione dei criteri utilizzati per la ricerca.

La soluzione individuata ha permesso di soddisfare a pieno quanto richiesto dal requisito [REQ4]. In particolare, si è introdotta, all'interno dell'ontologia di dominio, una tipologia specifica di proprietà che permette di selezionare i contenuti testuali in base ai quali ciascuna tipologia di risorsa viene indicizzata. Questo consente di avere un pieno controllo sui contenuti prodotti dalle ricerche ovvero quelli che diventano i punti di accesso al patrimonio informativo dell'archivio.

La semplice struttura dell'indice, basata su una tripla che associa direttamente la risorsa 'risultato' ed il testo che la descrive, permette di utilizzare il criterio di ricerca testuale in unione con altri criteri di restrizione. Il costo relativo all'esecuzione rimane infatti circoscritto alla giunzione derivante dai due predicati necessari all'utilizzo dell'indice (vedi listato 2.4) evitando di dover considerare le giunzioni relative al grafo unione di tutte le proprietà già incluse nell'indice.

Allo stesso tempo, la struttura realizzata non vincola l'implementazione delle funzionalità di ricerca agli strumenti offerti dal *repository* e dal relativo servizio SPARQL con cui il *dataset* viene gestito. Resta infatti possibile esportare le triple dell'indice in strumenti esterni che potranno utilizzare il contenuto del nodo *literal* per indicizzare la risorsa.

Dal punto di vista funzionale questa soluzione consente all'utente di ricercare i contenuti utilizzando criteri testuali liberamente scelti. Non viene richiesta una conoscenza pregressa del modello descrittivo e degli attributi di tipo testo previsti per ciascuna tipologia di entità. Sarà l'esperto di dominio che, selezionando a priori i contenuti dell'indice, farà sì che la ricerca sia circoscritta ai più significativi per ciascuna tipologia di entità del dominio.

L'aver dato una definizione formale del 'contesto informativo' in base al quale ciascun risultato viene selezionato in base ai criteri offre la possibilità di valutare sia l'efficacia dei criteri scelti in relazione agli

obiettivi della ricerca, sia la possibilità di avviare la consultazione partendo proprio dall'esplorazione del contesto.

Per contro, l'indice così realizzato non risulta selettivo, ovvero non permette di restringere la ricerca indicando il tipo di relazione tra testo e risultato cercato. Non risulta possibile specificare ad esempio che il termine "Roma" debba essere il luogo di prima rappresentazione di un'opera e non semplicemente far parte del testo che ne descrive l'ambientazione. Questo limite deriva dal fatto che si è scelto di non inserire nell'indice anche la relazione tra risorsa indicizzata e attributo testuale. Scelta motivata dalla necessità di lasciare svincolato il contenuto dell'indice dal modello che lo ha prodotto, rimanendo quindi un semplice contenuto descrittivo della risorsa.

3 Analisi dei contenuti

Quanto illustrato nel capitolo 2 mostra come sia possibile ottenere, attraverso l'adozione di ontologie formali, un modello concettuale omogeneo e coerente della storia del melodramma. Si è inoltre evidenziato come, attraverso la stratificazione dei livelli descrittivi, sia possibile far convivere una considerevole analiticità nella definizione semantica del dato con una efficace rappresentazione funzionale. Al tempo stesso, occorre sottolineare come la struttura reticolare del grafo renda possibile una navigazione circolare dei contenuti dell'archivio. Resta valida la tripartizione fondamentale del modello concettuale FRBR, ovvero la netta suddivisione tra il contenuto intellettuale o artistico di un'opera, la sua espressione storica e la relativa sedimentazione sul piano materiale. Opere, eventi e libretti restano quindi le dimensioni principali per accedere all'informazione del dominio.

Ciascuna di queste dimensioni offre la possibilità di esplorare la storia del melodramma da prospettive specifiche ed al tempo stesso funge da punto di ingresso verso gli altri piani di lettura. Si può partire dal repertorio di un librettista per studiare la fortuna della sua produzione seguendo il susseguirsi della cronologia delle rappresentazioni, o indagare la tradizione del testo di un libretto attraverso le diverse edizioni.

L'aver organizzato i contenuti dell'archivio in base ad un modello concettuale a semantica esplicita offre però la possibilità di indagare anche in altre direzioni. In particolare, si possono individuare ulteriori dimensioni di analisi in grado di sondare il patrimonio informativo lungo direttrici che offrano prospettive di lettura diverse.

L'obiettivo in questo caso è offrire allo studioso strumenti di accesso ai contenuti che facciano emergere nessi e relazioni tra le diverse entità del dominio. Il compito del sistema informativo sarà quindi quello di proporre all'utente elementi, indizi o suggerimenti, che possano essere utilizzati come spunti per un approfondimento dell'indagine. In questa prospettiva il sistema deve offrire delle dimensioni di analisi a supporto dello studioso, cui resta l'onere imprescindibile della verifica delle ipotesi formulate a partire dai risultati proposti dalle ricerche.

Nel capitolo vengono presentate tre dimensioni di analisi che sono state individuate come particolarmente efficaci nell'offrire prospettive di accesso al patrimonio informativo dell'archivio. A partire da queste verranno elaborate le strategie di ricerca e navigazione dei contenuti descritte nel capitolo 5.

3.1 Lo spazio ed il tempo

Spazio e tempo rappresentano le dimensioni fondamentali all'interno delle quali si colloca necessariamente ogni attività umana. Sono quindi il contesto imprescindibile per collocare correttamente le diverse entità oggetto di studio. Nel caso specifico delle arti performative, le relazioni che intercorrono tra l'evento e i testimoni documentali sono molteplici e articolate. La nozione di spazio e tempo in uno spettacolo operistico ha significati diversi. C'è lo spazio ed il tempo della rappresentazione di un'opera (la serata e il teatro in cui va in scena), il tempo e l'ambientazione dell'azione drammatica rappresentata sul palcoscenico e il tempo e lo spazio proprio dei documenti che dello spettacolo sono testimoni.

Lungi dal poter affrontare in questa sede una trattazione della definizione delle due dimensioni in relazione alla storiografia dello spettacolo, ci si limita a trattare le tipologie di informazioni relative a spazio e tempo presenti nell'archivio e a presentare alcuni casi d'uso significativi che verranno impiegati come riferimento nello sviluppo delle funzionalità di ricerca.

3.1.1 La dimensione geografica del melodramma

Per poter parlare di spazio in relazione allo spettacolo è necessario specificare quali elementi del fenomeno si vogliono descrivere ed analizzare. I dati che descrivono il rapporto tra atto performativo e spazio è infatti molteplice: si va dalla località in cui si trova il teatro come edificio che ospita la rappresentazione, lo spazio scenico inteso come contesto materiale in cui si svolge l'azione drammatica fino ad includere lo spazio ideale della narrazione⁶³. A corollario, vanno considerati tutti i luoghi relativi alle persone e ai documenti: dai riferimenti biografici a quelli bibliografici che testimoniano le vicende del processo creativo ed editoriale dell'opera. L'insieme di questi riferimenti alla dimensione geografica dello spettacolo sono associati nel modello descrittivo dell'archivio *Corago* alle entità principali: opera, evento e documento. Nel modello relazionale utilizzato dalla procedura di catalogazione questi vengono trattati come attributi testuali descrittivi della singola entità: manca quindi un indice generale delle località. Nella trasposizione in formato RDF, si è utilizzata la classe `crm:E44_Place_Appellation` per raccogliere i diversi toponimi censiti e poterli organizzare in un indice strutturato⁶⁴. In questo modo diventa possibile accedere ai contenuti dell'archivio a partire da chiavi comuni alle tre diverse sezioni dell'archivio.

3.1.1.1 Luoghi, teatri, ambientazioni

I riferimenti allo spazio nell'archivio sono di tre tipologie: località riferite a persone e documenti; teatri e luoghi in cui si sono tenuti gli spettacoli; luoghi di ambientazione delle opere. Le prime due tipologie sono trattate come campi testuali che riportano il toponimo o la denominazione dell'edificio. La terza compare all'interno delle didascalie relative all'ambientazione dell'opera riportata nei libretti e catalogata attraverso un metadato di tipo testuale.

I toponimi sono associati alle persone come luoghi di nascita e morte, alle opere come luoghi di prima esecuzione, ai documenti e riferimenti bibliografici come località di edizione e alle biblioteche come luoghi di residenza. Sono invece associati agli eventi le località (le 'piazze') e gli edifici teatrali in cui si è tenuto lo spettacolo. In entrambi i casi viene riportata la denominazione in forma di campo testuale. In fase di

⁶³ Per una trattazione generale del complesso rapporto tra teatro e spazio si rinvia a (Cruciani, 1993), per uno sguardo d'insieme ai problemi di storiografia degli spazi scenici e dei luoghi del teatro (Bignami e Azzaroni, 1997).

⁶⁴ Si rinvia al paragrafo 4.2.5 per la discussione della procedura di georeferenziazione dei toponimi.

trasposizione in RDF questi sono stati definiti attraverso la classe `crm:E19_Physical_object`. La denominazione è stata impiegata come discriminatore dell'identità del teatro ottenendo così un indice degli spazi teatrali⁶⁵.

L'informazione relativa alla relazione tra opera e luogo deve essere ricavata dalla descrizione dell'ambientazione non essendo esplicitata nel modello descrittivo dell'archivio⁶⁶. Ciascuna delle diverse relazioni del repertorio e degli spettacoli con lo spazio ha una sua specifica valenza dal punto di vista funzionale. L'unificazione delle diverse relazioni rispetto ad una comune dimensione d'analisi permette di accedere ai contenuti dell'archivio in funzione della loro distribuzione su base geografica in modo omogeneo.

3.1.1.2 Georeferenziazione

Per posizionare opere, eventi e documenti utilizzando una cartografia è però necessario associare all'indice dei toponimi le relative coordinate geografiche (o la relativa geometria, se si tratta di territori: regioni, paesi, continenti). Questa informazione, non presente nell'archivio *Corago*, viene ricavata da fonti informative esterne⁶⁷. L'uso delle coordinate geografiche all'interno dell'indice rende possibile sia rappresentare le entità che vi fanno riferimento sulla cartografia, ma anche predisporre una ricerca a partire dalla definizione sulla mappa di una specifica area di interesse.

3.1.2 La dimensione temporale

La dimensione temporale è essenziale per organizzare l'informazione di fatti storici. La loro collocazione lungo l'asse temporale è la preconditione per la ricostruzione del contesto storico in cui sono avvenuti e deve essere necessariamente coordinata alla loro collocazione spaziale. Come per quest'ultima si rende necessario ricondurre le diverse tipologie di entità dell'archivio ad una comune linea temporale (*timeline*). In questo l'utente può osservare lo svolgersi degli eventi e consultare i documentali collegati in una unica e coerente prospettiva diacronica.

Al tempo storico si aggiunge il 'tempo dello spettacolo', nella duplice accezione di momento della festa⁶⁸ e di tempo dell'agire scenico ovvero dello svolgersi dell'azione drammatica. In questo senso i due tempi vengono a sovrapporsi in modo dialettico: l'ambientazione temporale delle opere incontra il tempo storico nella percezione dello spettatore. Restituire il nesso tra tempo dell'azione drammatica e cronologia degli spettacoli offre una ulteriore prospettiva di lettura di questo rapporto.

I riferimenti cronologi presenti nell'archivio vengono definiti in base alla classe `crm:E52_Time-Span`. Di questi fanno parte le date di nascita e morte delle persone fisiche presenti nell'*authority file*; le date di composizione, di prima rappresentazione e di ambientazione del dramma; le date degli spettacoli e delle stagioni; le date di edizione dei documenti e della bibliografia.

⁶⁵ Si veda (Bonora, 2016, p. 48) per la possibilità di associare al teatro le relative coordinate geografiche attraverso l'allineamento con i Punti di Interesse (POI) della base di conoscenza LinkedGeoData.

⁶⁶ Si rinvia al paragrafo 4.2 per la discussione della metodologia e degli strumenti utilizzati a questo scopo.

⁶⁷ Si rinvia al paragrafo 4.2.5 per la descrizione della modalità di acquisizione delle coordinate geografiche a partire dai toponimi censiti nell'indice.

⁶⁸ «Il tempo dello spettacolo non è soltanto un tempo diverso [...] è il tempo ideale, il tempo della festa come celebrazione. [...] Entrando in sala, dunque, si compie il salto dal tempo reale al tempo della festa». Con questa chiave di lettura Fabrizio Cruciani propone di interpretare questo scarto come l'elemento centrale di quel momento fondante che è il Teatro del Rinascimento rispetto alle forme di spettacolo successive (Cruciani, 1987, p. 34).

Ciascuna di queste, con la relativa precisione⁶⁹, è confluita nell'unica *timeline* presente nel *dataset* RDF: lo scorrimento della *timeline* permette di ricavare la sequenza cronologia non solo degli eventi, ma anche delle entità ad essi associate.

3.1.3 Analisi coordinate spazio-tempo

L'unificazione dei riferimenti spazio-temporali delle entità presenti in repertorio, cronologia e documenti ha mira all'analisi coordinata di fenomeni in cui le due dimensioni sono un elemento essenziale (vedi [REQ5], paragrafo 1.6.3) come ad esempio l'analisi della diffusione di un'opera. Se consideriamo l'*Orfeo ed Euridice* di Christoph Willibald Gluck, a partire dalla sua prima rappresentazione a Vienna nel 1762, in archivio sono censite ventidue rappresentazioni dalla sua prima a Vienna nel 1762 fino al 1889 in tredici località diverse. L'unificazione delle coordinate e l'impiego della cartografia permette di rappresentare plasticamente il fenomeno della sua diffusione geografica in un arco temporale lungo oltre un secolo. Allo stesso tempo, è possibile analizzare il dato in modo sincronico. Ad esempio, nel 1762: oltre all'*Orfeo ed Euridice* di Gluck a Vienna, a Venezia debuttano *Il marchese villano* di Baldassarre Galuppi⁷⁰ e l'*Alessandro Severo* di Antonio Sacchini⁷¹.

Consultando l'archivio in prospettiva diacronica, è possibile studiare lo sviluppo della carriera di un interprete. Ad esempio, Carlo Broschi⁷², in arte Farinelli, risulta aver interpretato oltre sessanta ruoli tra il 1720 e il 1737. Nei primi anni la sua presenza si concentra sulle piazze di Roma e Napoli, con una prima apparizione a Londra nel 1724 per l'*Aquilio consolo*⁷³ di Francesco Bernardi. A partire dal 1730 lo troviamo nelle principali piazze italiane (da Firenze a Venezia) ed europee (Vienna e Monaco) per tornare quasi stabilmente a Londra tra il '35 e il '37. Complessivamente saranno Roma, Napoli e Venezia le piazze che registrano il maggior numero di presenze: undici ciascuna, con Londra che resterà invece la più frequentata in assoluto con ben sedici presenze.

Stesso approccio può essere utilizzato per studiare la diffusione di un'opera come la *Didone abbandonata*⁷⁴ di Pietro Metastasio. Rappresentata sei volte tra il 1742 e il 1769 tra Dresda, Napoli, Londra e Berlino; della quale sono censiti trentasette libretti in almeno quattro lingue (italiano, tedesco⁷⁵, inglese⁷⁶ e spagnolo⁷⁷).

Questi semplici esempi di lettura del patrimonio informativo contenuto nell'archivio sono resi possibili dalla armonizzazione delle due dimensioni ottenuta grazie all'astrazione dell'informazione relativa alle relazioni con lo spazio e il tempo del modello concettuale del CIDOC CRM.

⁶⁹ In funzione della precisione dell'informazione ricavabile dalla fonte in fase di catalogazione di ciascuna scheda di entità presente in archivio la precisione può variare dall'anno, mese e giorno (ad es. per gli spettacoli), al periodo dell'anno nel caso delle stagioni, fino al secolo o "periodo ideale" nel caso delle ambientazioni temporali delle opere.

⁷⁰ Riferimento URI dello spettacolo: <http://corago.unibo.it/resource/SPETTACOLI/7C00000925>.

⁷¹ Riferimento URI dello spettacolo: <http://corago.unibo.it/resource/SPETTACOLI/7C00000581>.

⁷² Riferimento URI dell'artista: <http://corago.unibo.it/resource/RESPONS/000005067400>.

⁷³ Riferimento URI dello spettacolo: <http://corago.unibo.it/resource/SPETTACOLI/0000401582>.

⁷⁴ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/0000297595>.

⁷⁵ Riferimento URI al libretto: <http://corago.unibo.it/resource/DOCUMENTI/0001369170>.

⁷⁶ Riferimento URI al libretto: <http://corago.unibo.it/resource/DOCUMENTI/DRT0013979>.

⁷⁷ Riferimento URI al libretto: <http://corago.unibo.it/resource/DOCUMENTI/0001279451>.

3.2 I personaggi come dimensione di analisi del testo drammatico

Lo studio della tradizione dei testi letterari del teatro d'opera costringe a confrontarsi con una varietà significativa di problematiche. Infatti, oltre che con le tradizionali difficoltà della filologia e bibliografia caratteristici del testo poetico è necessario affrontare anche i problemi tipici della filologia di testi a tradizione contaminata che sono soggetti a innovazione e rimaneggiamenti continui (Accorsi, 1989, p. 213). La forma in cui questi testi oggi vengono fruiti è frutto di una elaborazione strettamente connessa al processo di produzione e messa in scena dell'opera. Un processo produttivo che prende l'avvio con l'appalto della stagione all'impresario e che risponde ad un complesso sistema di convenzioni, prassi e rapporti tra committenza⁷⁸ e librettista.

Nel descrivere il lavoro del librettista Felice Romani, Alessandro Roccatagliati individua come premessa ineludibile per procedere alla definizione di soggetto e personaggi la conoscenza degli interpreti che saranno impiegati (Roccatagliati, 1996, p. 94). Ne consegue la necessità di mantenere una stretta corrispondenza tra caratteristiche del personaggio rappresentato e quelle del cantante che lo interpreta (*ibidem*, p. 88). Nel primo Ottocento il periodo utile per la stesura di un libretto era pari a circa quaranta giorni, dalla scelta del soggetto alla consegna al committente (*ibidem*, p. 67). Il processo compositivo partiva quindi dalla spoliatura di fonti preesistenti per individuare nuovi soggetti o riprendere testi già impiegati⁷⁹, il cui contenuto veniva successivamente adeguato, attraverso un lavoro a più mani, alle specifiche esigenze del nuovo allestimento⁸⁰. Nella stesura di un nuovo dramma diventa rilevante il forte nesso con gli antecedenti letterari e il profilo dei personaggi che questo determina⁸¹.

Nel teatro di prosa coevo, in cui il testo drammatico acquisisce un ruolo centrale nel nuovo sistema produttivo teatrale, l'interprete entra in relazione con il personaggio sullo schema del *carattere*⁸² (Jandelli, 2002, p. 64). La parte diventa quindi il materiale su cui, alla bisogna, si può intervenire in funzione delle esigenze sceniche o dell'interprete. Interventi che vengono fatti nel rispetto della rigida codifica dei ruoli rispetto a cui ciascuna parte deve trovare una collocazione (*ibidem*, p. 65). È così che nel passaggio dall'autore, al capocomico ed infine all'attore, il testo prima frammentato in parti, poi ricomposto sulla scena, assume una forma e identità autonome. Anche nel teatro di prosa il legame tra testo e personaggio sembra non sia dettato esclusivamente dalla sua funzione drammaturgica, ma anche da quanto i suoi tratti, le sue caratteristiche, risultino funzionali al processo produttivo che lo porterà in scena.

⁷⁸ L'adesione alle 'convenzioni teatrali', ovvero l'adattamento alla piazza ed alla compagnia per cui è stata commissionata l'opera è spesso determinante: «Se l'impresario dispone di due prime donne, per dire, si potrà, anzi si dovrà fare un dramma fondato su una rivalità erotico-politica femminile, poniamo *L'incoronazione di Poppea* o *Maria Stuarda*» (Bianconi, 2017, p. 198).

⁷⁹ «Il bricolage delle fonti è a volte capillare, *quasi mai dichiarato*» (Bianconi, 2017, p. 199) [corsivo mio].

⁸⁰ Cfr. (Bianconi, 2017, p. 198).

⁸¹ Nella prefazione alla *Fedra* andata in scena a Milano nel 1820 (URI del libretto: <http://corago.unibo.it/resource/DOCUMENTI/DMBM18072>) Luigi Romanelli, riferendosi alla sua *Fedra* di Padova dello stesso anno (URI dell'opera: <http://corago.unibo.it/resource/BRANI/0000097987>), afferma: «Lo stesso argomento fu da me trattato in un Melodramma, che si rappresentò in Padova nell'occasione dell'ultima fiera del Santo. Ma il presente Melodramma, fuorché la sostanza del fatto, nulla ha di comune col primo né per la condotta, né per la versificazione; imperciocché la diversità del Teatro, la qualità degli Attori, ed altre circostanze esigevano, che la composizione fosse del tutto nuova».

⁸² Ci si riferisce qui alla definizione di *carattere* come «l'insieme di tratti psicologici, morali e fisici che contraddistinguono un personaggio teatrale, cioè la sua identità e personalità scenica. Ma, specie nel primo Ottocento, si considerava il carattere un segno, un'impronta, una qualità costante e fissa (che poteva coincidere con una passione dominante) rigidamente codificata.» (Jandelli, 2002, p. Dizionario dei Ruoli, voce: Carattere).

Visto lo stretto rapporto tra personaggio e testo si è ritenuto opportuno utilizzare il personaggio come indicatore utile per individuare relazioni tra i testi delle diverse opere. In base a questa ipotesi la presenza di personaggi con lo stesso nome, o con caratteristiche quantomeno affini, può essere un indizio rilevante di una possibile relazione tra i due testi considerati. Si pone quindi il problema di identificare il personaggio non solo attraverso il nome, ma anche attraverso gli attributi che ne definiscono il profilo.

L'uso delle caratteristiche fisiche e morali come elementi in grado di definire il rapporto tra testo e personaggio ha avuto fortune alterne anche negli studi letterari. Già a partire da Ferdinand De Saussure il personaggio dei grandi cicli leggendari è un simbolo di cui possono variare nell'ordine il nome, le relazioni con gli altri personaggi, il carattere e le azioni (Stara, 2004, p. 185)⁸³. In Vladimir Propp, gli attributi del personaggio delle favole costituiscono la parte mutevole che riveste il nucleo funzionale stabile. Essi spiegano l'estrema varietà del genere a fronte di una sostanziale uniformità strutturale e tematica (Stara, 2004, p. 191). Se in Propp, gli attributi, le caratteristiche superficiali del personaggio sono quindi le 'grandezze variabili', a partire da Claude Lévi-Strauss questa distinzione viene superata ricomprendendoli all'interno della struttura narrativa del testo. Nelle successive analisi strutturaliste, a partire da Roland Barthes, sarà la nozione stessa di personaggio a risolversi in quella di soggetto agente, analizzato in base al suo ruolo nell'azione e non in relazione al suo essere (Stara, 2004, p. 195). D'altra parte, il personaggio si posiziona al centro del delicato equilibrio tra referenzialità, mimesi del reale e puro fantastico che rende coerente la 'logica' dell'universo drammaturgico disegnato dal testo. E' così che nel mondo di Don Giovanni lo spettatore è in grado di accettare tranquillamente che una statua presenti al dissoluto il conto dei suoi misfatti (Elam, 1988).

Nel libretto d'opera, i personaggi vengono presentati allo spettatore attraverso didascalie che corredano l'elenco delle *dramatis personae* con brevi descrizioni. In queste sono riportati in forma sintetica, e spesso intenzionalmente reticente, i tratti caratteristici del personaggio che danno la possibilità allo spettatore di identificarlo sulla scena e nello svolgersi dell'azione drammatica. Le descrizioni fornite dalle didascalie sono quindi una fonte utile a ricavare i caratteri salienti, o almeno ritenuti tali dall'estensore del libretto, per identificare e caratterizzare i personaggi di un'opera. Occorre incidentalmente precisare che la scelta di limitarsi in questo studio alle sole caratteristiche contenute nelle didascalie è di natura squisitamente pratica essendo le stesse già disponibili in formato digitale e direttamente associate al repertorio. Nulla vieta, e i risultati ottenuti anzi incoraggiano a farlo, di estendere la sperimentazione in futuro includendo i testi completi delle opere.

L'ipotesi dell'utilizzo della descrizione 'superficiale' delle caratteristiche del personaggio come indicatore della sua identità narrativa, e quindi come elemento in grado di fornire indicazioni circa il contenuto del testo drammatico, è strettamente legata alla possibilità di strutturare ed organizzarne il contenuto informativo con l'obiettivo di renderlo ricercabile.

A questo scopo è necessario in primo luogo definire un modello concettuale che ne organizzi i contenuti al fine di poter svolgere le diverse attività di analisi necessarie allo studio della tradizione dei testi. Un primo livello di analisi dovrà infatti permettere di identificare personaggi con descrizioni sovrapponibili ma nomi

⁸³ Ben prima degli studi novecenteschi, Alessandro Piccolomini, nella dedica ad Antonio Cocco de *La sfera del mondo* nell'edizione del 1561, tentava una classificazione dei personaggi della commedia in base alle loro caratteristiche, relazioni e vicissitudini. Da questa Daniele Seragnoli ricava uno schema concettuale delle caratteristiche fondamentali individuate dal Piccolomini articolate in: relazioni di parentela, condizione economica, età, professione, indole e stati d'animo (Seragnoli, 1987, p. 304). Si veda come in 4.2.1 per la sperimentazione dell'estrazione automatica delle caratteristiche dei personaggi vengano scelte proprio le relazioni di parentela.

diversi. Successivamente si dovranno aggregare le diverse istanze di personaggi che condividono la descrizione intorno ad una entità comune che ne rappresenti una sorta di capostipite. Raggiunto questo livello descrittivo avremo insiemi di personaggi omologhi per descrizione e caratteristiche ma con identità proprie.

Procedendo con l'analisi degli attributi riportati dalle descrizioni si potrà operare un confronto sulle singole caratteristiche e quindi sperimentare aggregazioni che prendano in considerazione solo un sottoinsieme di esse. Ad esempio, raggruppare tutti i personaggi che condividono un rango nobiliare o una specifica professione e hanno un legame parentale con un altro personaggio dell'opera. A questo livello si supera l'identità del personaggio andando a isolare le caratteristiche che lo qualificano come soggetto agente nell'azione drammatica. In altre parole, sono gli attributi a definire l'identità dell'archetipo e non più il nome.

Il modello concettuale del personaggio che è stato elaborato nasce con una specifica finalità: supportare lo studioso nell'individuazione di opere che presentino personaggi con caratteristiche simili tali da suggerire una origine comune. In questa prospettiva il personaggio viene considerato semplicemente come una dimensione di analisi funzionale allo studio delle relazioni tra i testi. La descrizione ottenibile, in considerazione della circoscritta funzione strumentale e del suo essere vincolata al limitato contenuto informativo delle didascalie, non potrà assumere valore di carattere generale. L'obiettivo non è quindi produrre un catalogo dei tipi drammatici del teatro d'opera, bensì una dimensione di analisi utile ad indagare il repertorio dei testi da cui derivano nel tentativo di individuare indizi che possano suggerire debiti o dipendenze che dovranno essere oggetto di analisi successive che travalicano gli obiettivi di questo studio.

L'efficacia del modello descrittivo proposto è stata valutata sulla base dei risultati presentati nel paragrafo 3.2.3 relativi all'applicazione di euristiche di analisi applicate ad un campione di opere in repertorio.

3.2.1 I contenuti dell'archivio sui personaggi

Le informazioni relative ai personaggi presenti in archivio sono, in larga parte, derivate dagli elenchi dei personaggi presenti nei libretti (vedi immagine 3.1). Di norma, il personaggio può essere descritto da uno a più attributi, per esempio: la professione svolta, il suo rango sociale, il luogo di origine o la provenienza, le relazioni che lo legano agli altri personaggi dell'opera, le caratteristiche d'animo o morali. Ad esempio, Cleonice⁸⁴ sarà la "regina di Siria, amante corrisposta d'Alceste", oppure Tirso⁸⁵ il "villano affittuale de' beni d'Euristene in villa".

In alcuni casi viene data qualche ulteriore informazione utile a contestualizzare il personaggio nell'azione drammatica come avviene per Semiramide⁸⁶ che appare "in abito virile sotto nome di Nino re degli Assiri, amante di Scitalce, conosciuto ed amato da lei antecedentemente nella corte d'Egitto come Idreno".

⁸⁴ Cleonice personaggio nel "Demetrio, re di Siria", libretto di Pietro Metastasio, compositore Gioacchino Cocchi andata in scena al King's Theatre in the Haymarket, Londra, il 08/11/1757 (Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/0000050181>).

⁸⁵ Tirso compare in "Demetrio tiranno", libretto di Aurelio Aureli, compositore Bernardo Sabatini andato in scena al Teatro Nuovo di Piacenza nel 1694 (URI: <http://corago.unibo.it/resource/BRANI/0000050215>).

⁸⁶ Personaggio principale della "Semiramide riconosciuta" di Pietro Metastasio, compositore Leonardo Vinci che ha debuttato a Roma al Teatro delle Dame il 06/02/1729 (URI: <http://corago.unibo.it/resource/BRANI/0000051966>).



Immagine 3.1: Pagina del libretto con l'elenco dei personaggi della Semiramide riconosciuta⁸⁷

In altri casi le informazioni sono strettamente legate al contesto in cui vengono presentate, così è per Fausta che è semplicemente indicata come “sua moglie”, in riferimento a Costantino, nell’omonimo libretto di Giovanni Albinoni⁸⁸.

Le descrizioni, espresse in questa varietà di forma e contenuto, costituiscono il materiale da cui si è partiti per ricavare una descrizione strutturata del personaggio. A questo scopo si è ipotizzato di segmentare i contenuti della descrizione in unità informative elementari che specificassero le caratteristiche atomiche del personaggio, ad esempio: la professione, l’età o il rango sociale. La segmentazione del testo è stata operata su base morfosintattica ed ottenuta attraverso l’impiego di tecniche di analisi del linguaggio naturale (*Natural Language Processing* - NLP). Le diverse metodologie e gli strumenti che sono stati sperimentati per la segmentazione dei testi e l’estrazione dell’informazione confluita nel modello concettuale della ‘dimensione del personaggio’ vengono discusse nel paragrafo 4.2. Nel paragrafo successivo 3.2.2 viene invece presentato il modello concettuale elaborato per strutturare la dimensione ‘personaggio d’opera’ e l’articolazione nei diversi livelli utilizzati per la successiva analisi delle fonti letterarie. Sulla base della rappresentazione ottenuta in base al modello concettuale, sono state elaborate le euristiche di analisi delle dipendenze tra i testi letterari discusse nel paragrafo 3.2.3.

3.2.2 Il modello descrittivo del personaggio d’opera

Per poter utilizzare le descrizioni dei personaggi come chiavi di accesso ai testi è stato necessario definire un modello concettuale in grado di rappresentarne le caratteristiche in modo strutturato.

⁸⁷ Copia digitale integrale del libretto reperibile all’indirizzo:
<http://corago.unibo.it/esemplare/7AE0002320/DD00054443>.

⁸⁸ *Costantino*, libretto di Giovanni Kreglianovich Albinoni, compositore Joseph Hartmann Stuntz, rappresentata a Venezia l’8/02/1820 presso il Teatro La Fenice.

Il modello deve rispondere a due esigenze principali: fornire una rappresentazione analitica delle informazioni provenienti dalle descrizioni e permettere una loro generalizzazione in livelli descrittivi organizzati gerarchicamente.

Partendo dalle descrizioni sintetiche presenti nelle didascalie, le informazioni che qualificano il personaggio devono essere organizzate per consentire confronti tesi ad individuarne similarità o differenze. I contenuti testuali delle descrizioni devono essere segmentati in unità informative che rappresentino le caratteristiche elementari del personaggio descritto. Queste unità possono essere considerate gli attributi elementari che qualificano il singolo personaggio. Sarà quindi possibile confrontare i personaggi semplicemente confrontando l'elenco dei loro attributi elementari.

Ciascuna caratteristica espressa da un attributo potrà essere ulteriormente segmentata e sintetizzata. Ad esempio, un primo livello di segmentazione della descrizione di Cleonice “regina di Siria, amante corrisposta d'Alceste” potrà essere: “regina di Siria” e “amante d'Alceste”. Un successivo livello può qualificarla semplicemente come “regina” o “amante”.

In funzione del grado di analiticità della segmentazione delle descrizioni, i confronti tra personaggi possono essere effettuati a livelli progressivi di generalizzazione. Ad esempio, si possono confrontare personaggi che condividono la professione o il rango sociale ma differiscono per età; oppure, svolgono mestieri diversi ma hanno gli stessi legami di parentela con altri personaggi dell'opera. Generalizzando ulteriormente si può voler confrontare tutti i personaggi appartenenti a ranghi militari o legati da specifici rapporti di parentela.

La segmentazione della descrizione in attributi puntuali risponde alla necessità di individuare un primo livello di aggregazione dei personaggi che condividono le medesime caratteristiche comuni. Questo primo livello di aggregazione è rappresentato nel modello concettuale dall'entità ‘figura drammatica’ (vedi paragrafo 3.2.2.1).

La figura drammatica è definibile come il personaggio capostipite di un insieme di personaggi che, sebbene compaiono in opere diverse, possono essere considerati omologhi in base alle loro descrizioni. Una figura drammatica rappresenta quindi un gruppo di personaggi che condividono l'intero insieme di attributi elementari derivati dalle rispettive didascalie. Un esempio di questo tipo di aggregazione basata sull'analisi delle descrizioni si può avere considerando il personaggio di “Micco Papera” che, con questa denominazione, compare in tre diverse opere.

La prima è *Il fantastico*⁸⁹, libretto di Gennaro Antonio Federico, compositore Leonardo Leo, andata in scena nella stagione di carnevale del 1743 al Teatro Nuovo di Napoli. La didascalia del libretto lo descrive come personaggio “che fa da cavaliere errante col nome di D. Battinferno, detto il Cavaliere della Malasciagura”. Compare poi in *Il nuovo Don Chisciotte*⁹⁰, libretto dello stesso Federico e riveduto da Antonio Palomba, andato in scena nell'autunno del 1748 al Teatro dei Fiorentini di Napoli con le musiche di Leonardo Leo e Pietro Comes. Qui lo troviamo descritto come “napoletano d'umor sciocco e stravolto, che fa da Cavaliere errante col nome di Don Battinferno”. Troviamo nuovamente un personaggio con lo stesso nome in *Il ritratto*⁹¹, un libretto di Francesco Saverio Zini, messo in musica da Francesco Bianchi e andato in scena a Napoli, al Teatro Nuovo, nel settembre 1787. Viene indicato come “rustico negoziante di Sessa il quale, avendo sposata Chiaretta, la conduce alla di lui patria”.

⁸⁹ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/0000359090>.

⁹⁰ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/0000368723>.

⁹¹ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/7A00333599>.

Anche da una lettura superficiale delle descrizioni risulta evidente la vicinanza tra i primi due “Micco Papera” e la distanza di questi dal terzo. La relazione tra i due personaggi è confermata dalla relazione tra i due testi drammatici, attestata anche in letteratura⁹²: *Il nuovo Don Chisciotte* risulta tratto da *Il fantastico*. In questo esempio i primi due personaggi, qualificati con gli attributi di “cavaliere errante” e “col nome di Don Battiferro”, possono essere ricondotti ad una comune ‘figura drammatica’. Il terzo Micco Papera, caratterizzato dall’essere “un rustico negoziante di Sessa” e “marito di Chiaretta” dovrà essere invece associato ad una ‘figura drammatica’ distinta. Volendo generalizzare ulteriormente, i primi due, evidentemente debitori verso la figura letteraria di Don Chisciotte de la Mancia, potranno essere annoverati nella categoria archetipica dei “cavalieri erranti”, il terzo nella più prosaica categoria dei “mariti negozianti”.

I casi di personaggi che derivano dalla mitologia classica come Enea, pongono un ulteriore spunto di riflessione su come sia opportuno articolare la descrizione del personaggio d’opera quando, all’identità derivata dalle caratteristiche proprie, si aggiungano elementi esterni derivanti dalla tradizione letteraria del personaggio storico considerato.

Un personaggio con il nome di Enea è attestato in *Corago* in oltre centoventi opere datate tra il 1628 e il 1875. Lo vediamo rappresentato nelle diverse fasi del suo mito: la battaglia di Troia⁹³; il soggiorno in Cartagine e la relazione con Didone⁹⁴; l’approdo in Italia e le nozze con Lavinia⁹⁵. Ai diversi momenti delle vicende occorse ad Enea corrispondono altrettante descrizioni: a Troia è indicato come “marito di Creusa”⁹⁶, in Italia è un “prencipe de’ troiani”⁹⁷. In tutti i casi il riferimento è alla figura di Enea ed al suo mito. Si pone quindi la necessità di introdurre un livello descrittivo che consenta di rappresentare l’identità di un personaggio mitologico o storico a prescindere dalle diverse caratteristiche che lo descrivono nelle fonti. Nel modello concettuale proposto viene definita ‘identità narrativa’.

Questo riferimento ad una comune ‘identità narrativa’ di più figure drammatiche si arricchisce di ulteriori elementi quando viene messo in scena un personaggio storico. Il personaggio “Giulio Cesare” è attestato con questa forma del nome in *Corago* in almeno ventiquattro opere⁹⁸. Le opere traggono spunto dagli episodi salienti della sua vicenda storica: dalle campagne militari nelle Gallie⁹⁹ all’Egitto¹⁰⁰, fino alla tragica fine a Roma¹⁰¹. Nel caso di Cesare, l’identità narrativa comune ai diversi personaggi deve necessariamente confrontarsi con la figura del personaggio storico nella doppia prospettiva dello spettatore contemporaneo dell’opera e in prospettiva diacronica.

⁹² Vedi (Sartori, 1990) e (Cotticelli e Maione, 2016).

⁹³ In *Il greco in Troia*, libretto di Matteo Noris, messo in scena a Firenze nel 1689. Riferimento URI dell’opera: <http://corago.unibo.it/resource/BRANI/APC0003468>.

⁹⁴ Ad esempio, in *Didone*, opera composta da Francesco Cavalli su libretto di Giovanni Francesco Busenello, rappresentata in Venezia nel 1641. Riferimento URI dell’opera: <http://corago.unibo.it/resource/BRANI/APC0004044>.

⁹⁵ In *Le nozze d’Enea con Lavinia*, opera di Claudio Monteverdi su libretto d’anonimo, messa in scena a Venezia nel 1641. Riferimento URI dell’opera: <http://corago.unibo.it/resource/BRANI/APC0004359>.

⁹⁶ Riferimento URI alla scheda del personaggio: <http://corago.unibo.it/resource/PERSONAG/0000628146>.

⁹⁷ Riferimento URI alla scheda del personaggio: <http://corago.unibo.it/resource/PERSONAG/0000652779>.

⁹⁸ Mentre un personaggio denominato semplicemente “Cesare” compare in almeno settantasei opere.

⁹⁹ In *Giulio Cesare nelle Gallie*, libretto di Michelangelo Prunetti, compositore Giuseppe Nicolini, messa in scena a Roma nel 1819, viene definito “supremo comandante dell’Esercito romano”. Riferimento URI dell’opera: <http://corago.unibo.it/resource/BRANI/0000097859>.

¹⁰⁰ Tra le molte, la prima in ordine cronologico è *Giulio Cesare in Egitto*, librettista Giacomo Francesco Bussani, compositore Antonio Sartorio, messa in scena a Venezia nel 1676, in cui è indicato come “primo imperator de’ Romani”.

¹⁰¹ Tra le varie, *La morte di Cesare*, su libretto di Gaetano Sertor e musiche di Francesco Bianchi, rappresentata in Venezia nel 1788, in cui è definito “dittatore”.

Il modello concettuale deve quindi prevedere tre distinti livelli di astrazione: il primo dedicato a strutturare la descrizione del personaggio derivata dal contenuto testuale delle didascalie; il secondo deve permettere l'analisi delle caratteristiche comuni a più personaggi associandole ad una entità capostipite del gruppo; al terzo livello le diverse entità potranno a loro volta essere aggregate ed associate a forme archetipiche astratte identificate esclusivamente in base alle caratteristiche, o a identità narrative che facciano riferimento ad un mito o ad un personaggio storico.

Al primo livello il modello tratterà ancora personaggi intesi come elementi della *expression* dell'opera e legati al dato storico della rappresentazione: il materiale informativo di questo livello è costituito esclusivamente dalla superficie lessicale delle didascalie così come derivata dai libretti. Si avrà una diversa segmentazione dei contenuti e una organizzazione degli stessi orientata all'individuazione di caratteristiche comuni. Al secondo livello vengono prese in considerazione gli insiemi omogenei di caratteristiche a cui viene assegnata una denominazione 'collettiva': la figura drammatica. Infine, al terzo livello verranno rappresentati solo gruppi di attributi che formano archetipi astratti di caratteristiche o identità che fanno riferimento ad un contesto esterno al repertorio.

3.2.2.1 Le entità e le relazioni del modello.

Il modello concettuale risultante è stato elaborato con l'obiettivo di organizzare le informazioni presenti nell'archivio in modo da sperimentare il personaggio come dimensione di analisi del repertorio dei testi drammatici. Sono state definite le diverse articolazioni che rappresentano il personaggio nei tre livelli descrittivi ipotizzati come necessari per poter applicare le euristiche di analisi delle dipendenze tra i testi illustrate nel paragrafo 3.2.3.

Le definizioni delle entità previste dal modello sono:

Personaggio: soggetto citato in fonte come parte dell'azione drammatica. Rappresentata dalla classe frbroo:F38_Character.

Attributo: termine, sintagma o predicato derivato dalla descrizione del personaggio che ne specifica una determinata caratteristica. Rappresentata dalla classe corago:C17_Character_Feature.

Figura Drammatica: figura rappresentativa di uno o più personaggi omologhi, ovvero sovrapponibili in termini di caratteristiche e ruolo svolto nelle rispettive azioni drammatiche, che compaiono in opere diverse. Rappresentata dalla classe corago:C22_Dramatic_Figure.

Archetipo: insieme di attributi caratterizzanti un tipo di figura drammatica. Rappresentata dalla classe corago: C23_Archetype.

Identità Narrativa: soggetto dotato di una propria identità e collocazione spazio-temporale precisa che viene rappresentato in momenti della propria esistenza o con attributi diversi da una o più figure drammatiche che risultano far riferimento ad esso come entità unica ed autonoma. Rappresentata dalla classe corago:C24_Narrative_Identity.

Personaggio Storico: soggetto o entità attestata storicamente cui una identità narrativa o un archetipo fanno riferimento. Rappresentata dalla classe corago: C25_Historical_Figure.

Le relazioni che intercorrono tra le entità del modello sono rappresentate nel diagramma 3.1.

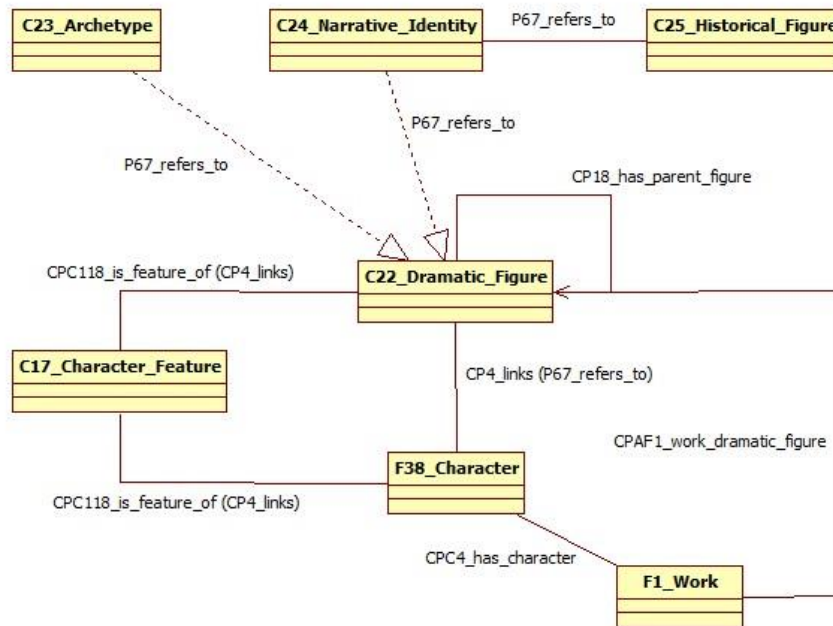


Diagramma 3.1: Il modello descrittivo del personaggio d'opera

In particolare occorre segnalare l'impiego di due livelli di specializzazione della proprietà di base `crm:P67_refers_to` e della proprietà equivalente definita nel Corago SM: la `corago:CP4_links`.

Il primo livello è definito a livello di navigazione attraverso la proprietà `corago:CPC118_is_feature_of`, il secondo livello è utilizzato per annotare gli esiti in fase di analisi con `corago:CPAF1_work_dramatic_figure` che rappresenta la relazione diretta individuata tra un'opera e le figure drammatiche associate.

Ad integrazione del modello descrittivo del personaggio, nell'ontologia di dominio, sono state definite due proprietà utilizzate per l'indicizzazione dei contenuti testuali delle descrizioni. In modo speculare alla definizione della proprietà astratta `corago:CPA5_ft_indexed_by`, è prevista una proprietà `CPA6_an_indexed_by` che consente di indicizzare le singole figure drammatiche. L'indicizzazione avviene in base ai testi individuati da specifiche proprietà figlie della `corago:CPA7_analytical_feature`. Il meccanismo ricalca quello descritto nel paragrafo 2.6.3.3 utilizzato per costruire l'indice full-text del sistema.

3.2.2.2 Come si popolano e si gestiscono i contenuti: aggregazione in base alle figure drammatiche

I primi due livelli descrittivi previsti dal modello sono stati popolati attraverso una serie di operazioni di normalizzazione e aggregazione dei nomi e delle descrizioni dei personaggi presenti in archivio. L'obiettivo di queste attività è consentire l'individuazione di gruppi di personaggi riconducibili ad una unica figura drammatica selezionata come capostipite.

Il procedimento utilizzato per attestare un personaggio alla figura drammatica di riferimento è articolato in quattro fasi:

1. normalizzazione dell'ortografia dei nomi e delle descrizioni;
2. raggruppamento (*clustering*) dei personaggi a parità di descrizione;
3. reificazione del personaggio capostipite a figura drammatica di riferimento;
4. analisi critica delle figure drammatiche per l'individuazione della figura capostipite.

Nella prima fase si è proceduto ad una verifica ortografica dei termini presenti nelle descrizioni e nelle forme dei nomi per risolvere errori in fase di catalogazione e oscillazioni ortografiche che avrebbero introdotto rumore in fase di applicazione del criterio di confronto.

Nella seconda fase è stato applicato il criterio di raggruppamento basato sulla comparazione dei nomi e delle descrizioni, ovvero: due personaggi sono stati considerati accorpabili se la loro descrizione coincide, al netto delle operazioni di normalizzazione.

Nel terzo passaggio, ottenuti i raggruppamenti di personaggi (*cluster*) se ne è selezionato uno che è stato assunto come figura drammatica di riferimento dell'intero gruppo. Il capostipite è stato selezionato su base cronologica, ovvero si è individuato il personaggio all'interno del gruppo che compare nell'opera più antica (antecedente della serie).

Col quarto passaggio si è proceduto a campione a verificare l'efficacia del *clustering* e l'eventuale inserimento nel gruppo di personaggi la cui denominazione o descrizione presentasse spiccate similarità, non rilevate nel secondo passaggio.

Il processo è stato supportato da una procedura sviluppata come estensione della procedura tradizionale di catalogazione impiegata dal progetto Corago, soluzione adottata in funzione della necessità di operare correzioni sulle schede descrittive di opere e personaggi. Il risultato è stato estratto e trasposto in formato RDF in base al modello descritto in 3.3.2.1 e riversato nel *dataset* generale.

Il risultato ottenuto ha permesso di popolare il modello fino al secondo livello: l'anagrafica delle 'figure drammatiche'. Il livello di astrazione successivo, che consiste nell'individuazione di archetipi o identità narrative, richiede una analisi dei contenuti delle descrizioni in grado di estrarre le caratteristiche peculiari della figura drammatica e la costruzione di un dizionario degli attributi significativi. Ad esempio, caratteristiche riferite all'individuo come l'età, il sesso, la professione o il rango sociale, ma anche alle relazioni che il personaggio ad altri personaggi all'interno dell'opera come i rapporti di parentela o di relazione: dall'amicizia ai vincoli di servitù.

Se si considera, la figura di Don Giovanni, assunta ad archetipo, questa deve essere qualificata quantomeno come uomo, aristocratico, scapolo ma con numerose amanti (Macchia, 1966), (Risé, 2006). Questi attributi, che definiscono i tratti identitari dell'archetipo, devono essere estratti dalle descrizioni. Si tratta quindi di interpretare il contenuto testuale della descrizione del personaggio per derivarne un sottoinsieme il cui valore semantico, espresso formalmente attraverso un vocabolario, confluisca in una descrizione omogenea di attributi qualificanti l'archetipo.

Questo tipo di analisi deve essere condotta dall'esperto di dominio attraverso una lettura critica del testo. Nel capitolo 4 viene presentata la sperimentazione dell'estrazione di informazione da contenuti espressi in linguaggio naturale mediante procedure parzialmente supervisionate (vedi paragrafo 4.2.1).

3.2.3 Analisi della tradizione dei testi attraverso i personaggi

L'ipotesi in base alla quale il personaggio può essere utilizzato come indicatore di relazioni che intercorrono tra i testi letterari del melodramma è stata verificata utilizzando il primo livello del modello descrittivo del personaggio d'opera. Il modello definito e popolato come illustrato nel paragrafo 3.2.2 è stato utilizzato per effettuare le ricerche comparative¹⁰². L'approccio utilizzato per l'indagine delle possibili corrispondenze si basa su euristiche che evidenziano parziali sovrapposizioni nelle compagini di personaggi tra due o più opere.

¹⁰² Si rinvia al paragrafo 5.2.2 per la descrizione degli strumenti realizzati a supporto delle ricerche e le relative modalità operative.

Ad esempio, il libretto de *Lo spirito di contraddizione*¹⁰³, di Caterino Mazzolà, andato in scena nel 1785, ha in comune con *L'ubbidienza per astuzia* dello stesso Mazzolà (1798) cinque personaggi con nomi e descrizioni sostanzialmente sovrapponibili (vedi tabella 3.1).

Tabella 3.1: Esempio di comparazione tra elenchi di personaggi

<i>Lo spirito di contraddizione (1785)</i>		<i>L'ubbidienza per astuzia (1798)</i>	
Nome	Descrizione	Nome	Descrizione
il dottore Oppositi	medico, marito di Dorimene	il Dottore	medico e speciale, marito di Dorimene
Dorimene	n.d.	Dorimene	sua seconda moglie
Lucilla	nipote del Dottore, amante di Lindoro	Lucilla	amante di Lindoro
don Pandolfo	ammalato immaginario in casa del Dottore	don Pandolfo	ammalato immaginario in casa del Dottore
Nespola	servitore	Nespola	servitore
		Lindoro ¹⁰⁴	giovane in pratica di medicina alloggiato in casa del Dottore

Una ricerca di tipo *full-text* sull'indice dei personaggi e delle relative descrizioni consente di individuare le opere sulla base di due criteri elementari:

- 1) nomi personaggi corrispondenti;
- 2) similarità o coincidenza della didascalia (anche a prescindere dai nomi).

Una volta individuate le opere candidate, la verifica dei testi ed il giudizio del musicologo sulla validità dell'indizio può convalidare o meno la sussistenza di una relazione tra i testi esaminati¹⁰⁵. L'analisi critica ha consentito di considerare *L'ubbidienza per astuzia* [1798] come una sostanziosa revisione di *Lo spirito di contraddizione* [1785]: riduzione dai due atti ad atto unico; ripresa di porzioni rilevanti della versificazione di [1785], presenza di numerose parti nuove. In conclusione, tra le due opere può essere stabilita una relazione di rielaborazione e revisione ad opera dello stesso autore.

¹⁰³ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/7A00336580>.

¹⁰⁴ L'elenco qui riportato non segue l'ordine di elencazione dei personaggi ricavabile dalla fonte dove Lindoro compare come quarto. Si è preferito evidenziare l'allineamento con i personaggi corrispondenti dell'opera comparata (nel cui elenco Lindoro non è riportato).

¹⁰⁵ Di norma l'indagine testuale è stata condotta a partire dalle prime scene del primo atto. Indicativamente le prime cinque o sei scene; in casi particolari l'analisi è stata estesa ad ulteriori scene del primo atto o degli atti successivi.

Utilizzando questa metodologia di analisi si è indagato il repertorio tra 1700 e 1800 individuando oltre 200 nuove relazioni tra testi di cui 188 verificate e riportate all'interno dell'archivio¹⁰⁶. Alcuni dei casi più significativi per le diverse tipologie di relazione vengono riportati nel paragrafo 3.3.3.1. L'insieme dei casi analizzati è pubblicato come *dataset* all'indirizzo: <https://doi.org/10.6084/m9.figshare.9756788> (verificato il 15/01/2020).

3.2.3.1 Risultati: relazioni tra libretti identificate attraverso la ricerca di figure drammatiche simili.

I risultati ottenuti sono organizzati in un *dataset* rappresentato in forma testuale con i casi riportati in schede articolate su quattro voci: opere individuate, criteri di interrogazione ed indizi di similarità rilevati, analisi critica e tipologia di relazione individuata.

Per ciascun caso vengono riportati i seguenti elementi:

- 1) **Opere individuate:** vengono indicati gli antecedenti cronologici ed i testi derivati.
- 2) **Criteri e indizi:** vengono specificati i criteri utilizzati per la ricerca e gli indizi individuati.
- 3) **Analisi critica:** viene riportato l'esito della analisi critica effettuata sui testi individuati.
- 4) **Tipo di relazione:** viene indicata la tipologia di relazione individuata tra l'antecedente ed il derivato.

Le tipologie di relazione tra antecedente e testi derivati sono le seguenti:

1. **nuovo:** testo originale;
2. **ripresa di:** nel libretto 'figlio' il dramma (escluse le arie) risulta sostanzialmente identico o presenta varianti minime rispetto all'originale;
3. **revisione di:** nel libretto 'figlio' il dramma (escluse le arie) presenta interventi di una certa entità nel dettato e nell'ossatura drammatica¹⁰⁷ rispetto al testo 'padre' (cancellazione di scene; scene aggiunte; scene modificate), ma conserva ampie porzioni del dettato del 'padre';
4. **rielaborazione di:** nel libretto 'figlio' il dramma (escluse le arie) si discosta in maniera notevole dal libretto 'padre', sia nell'ossatura sia nel dettato, e li ricalca soltanto a tratti;
5. **rifacimento di:** tessitura sostanzialmente nuova del testo drammatico (escluse le arie) rispetto al testo 'padre', ma con osservanza più o meno aderente dell'intreccio e occasionali ripescaggi di versi;
6. **da accertare:** non è stata effettuata la comparazione dei testi.

Si riportano di seguito alcuni casi esemplificativi delle relazioni individuabili.

Caso:1	Opere individuate
	<i>L'incanto superato</i> , favola romanzesca per musica, Süßmayr-Bertati, Vienna, 1793 (http://corago.unibo.it/resource/BRANI/0000194616)
	<i>Melinda</i> , favola romanzesca in musica, Nasolini-Bertati, Venezia, 1798 (http://corago.unibo.it/resource/BRANI/7A00313413)
	Criteri utilizzati ed indizi ricavati

¹⁰⁶ Tra questi in una decina di casi si è riscontrata la necessità di correggere od integrare l'attribuzione di testi indicata nel *Grove Music Online*. Ad esempio: *I vecchi delusi* di G. Marinelli e G. Palomba, 1793 ed *I vecchi burlati*, G. Marinelli e G. Palomba, 1795 in (Libby and McClymonds, 2001) sono elencate come opere diverse. In realtà [1795] contiene modifiche di piccola entità nei recitativi ed alcune arie sostituite rispetto a [1793]: si tratta con ogni probabilità della stessa opera.

¹⁰⁷ Per i concetti di 'ossatura del dramma', 'entità del contenuto' e 'dettato del testo' si rimanda a (Bianconi et al., 2018).

	L'elenco dei personaggi presenta significative affinità; quattro coincidono, tre modificati: Falsirena, fata / Melinda, fata; Oliviero=; Logistilla/Malagigi; Lidia=; Erminia =; Carotta=; Alcuino/Seriprando.
	Analisi critica
	L'impianto del plot sembra lo stesso: il cavaliere approda sull'isola della Fata ...; testo completamente riscritto e modificato in parte. I personaggi Logistilla/Malagigi hanno nomi diversi (Logistilla ha il nome della fata sorella di Alcina e Morgana), ma rivestono lo stesso ruolo drammaturgico in relazione al cavaliere Oliviero. Questo caso può essere usato come esempio per distinguere il concetto di figura drammatica (Logistilla, il nome riconduce alla fata capotistipite) e quello di archetipo (Logistilla/Malagigi, sebbene di nome e sesso diverso rivestono la stessa funzione drammaturgica in relazione ad Oliviero).
	Tipo di relazione
	Rielaborazione

Caso:2	Opere individuate
	<i>Le vipere han beccato i ciarlatani</i> , Mosca-Anonimo, Torino, 1801 (http://corago.unibo.it/resource/BRANI/0000096958)
	<i>Non essere geloso</i> , Pilotti-Anonimo, Bologna, 1815 (http://corago.unibo.it/resource/BRANI/0000104664)
	Criteri utilizzati ed indizi ricavati
	Affinità nell'elenco dei personaggi; tre coincidono, quattro sono modificati: Bellisle/Belille; Gianfrullo=; Carotta=; Teresina=; Rosina/Carlotta; La-Bore/Laborde; Nardo/Tonino.
	Analisi critica
	Affinità nell'impianto del plot; versificazione completamente diversa.
	Tipo di relazione
	Rifacimento

Caso:3	Opere individuate
	<i>L' amante ridicolo</i> , Piccinni-Pioli, Napoli, 1757 (http://corago.unibo.it/resource/BRANI/0000191648)
	<i>Il conte Baggiano</i> , Piccinni+AA.VV.-anon., Roma, 1770 (http://corago.unibo.it/resource/BRANI/0000180558)

	Criteri utilizzati ed indizi ricavati
	Elenco dei personaggi: non è disponibile la riproduzione della 1° ediz. (Roma 1762); nell'edizione Bologna 1763 c'è un personaggio in più (Pasquino), in Lisbona 1763 (stesso titolo di Roma) i personaggi sono quattro; tra Bologna e Lisbona ci sono alcune piccole divergenze; diverse soltanto alcune arie, ma rispettata la scansione delle scene.
	Analisi critica
	[1770] sembra una versione riveduta di [1763]: identico il recitativo iniziale; poi arie per lo più nuove, ma riprese almeno due; plot sostanzialmente inalterato; recitativo modificato largamente. Grove elenca [1770] tra le opere dubbie di Piccinni; credo che l'opera possa essere ragionevolmente attribuibile a Piccinni e può essere ritenuta attendibile la dichiarazione del lib. 1770 che indica l'autore in Piccinni con l'aggiunta di arie nuove di altri autori, contrassegnate da asterisco secondo la prassi corrente. Sempre in riferimento a L'amante ridicolo di Piccinni-Pioli, è stato possibile identificare anche una partitura manoscritta di Piccinni dello stesso intermezzo ma con un diverso titolo, e pertanto considerata un'altra opera di Piccinni.
	Tipo di relazione
	Revisione

Caso:4	Opere individuate
	<i>La fede ne' tradimenti</i> , Fabbrini-Gigli, Siena, 1689 (http://corago.unibo.it/resource/DOCUMENTI/DRT0017805)
	<i>La Climene</i> , Trento-Caravita, Londra, 1811 (http://corago.unibo.it/resource/DOCUMENTI/0001418610)
	Criteri utilizzati ed indizi ricavati
	Sono sovrapponibili i nomi dei personaggi: Garzia=Garcia (re di Navarra), Fernando=Ferdinand (conte di Castiglia), Elvira (sorella del conte di Castiglia). Differisce il nome della sorella di Garzia: Anagilda in [1689], Climene in [1811].
	Analisi critica
	[1811] appare un rifacimento di [1689]: libretto completamente nuovo, riprende solo alcuni versi da 1689. Esempi: In I,3: [1689] "Gar. Qual torbido pensiero [...]", [1811] "Gar. Qual torbido pensier t'ingromba l'alma?". In I,5: [1689] "Ana. Garzia, questo è il Tesoro, che riserbo al mio sposo, ed è come vedrai, al nostro Genitor costato assai.", [1811] "Cli. Eccolo, è qui il tesoro, che riserbo al mio sposo, ed è come saprai, al nostro Genitor costato assai."

	[1689] in I,5: “Ana. [...] <i>mostra una spoglia insanguinata e tagliata</i> . Questo lacero ammanto, che nel sangue del Padre intriso è tutto”; [1811] in I,3 “Gar. <i>scopre il bacile e vede una veste insanguinata e tagliata</i> . Questo lacero ammanto, che nel sangue del padre intriso è tutto”.
	Tipo di relazione
	Rifacimento

Caso:5	Opere individuate
	<i>Il geloso ravveduto, o sia I pazzi</i> , De Blasis-?, Napoli, 1784 (http://corago.unibo.it/resource/BRANI/7A00337322)
	<i>Il geloso corretto</i> , Gnecco-Camagna, Venezia, 1803 (http://corago.unibo.it/resource/BRANI/0000097330)
	Criteri utilizzati
	Quattro personaggi modificati, uno coincidente ed uno mancante: Margherita Fajensara/Lucietta; Agata/Beatrice; Mauriello/Gregorio; Giorgio=; Panfilio=NO; Camillo/Federico.
	Analisi critica
	Sostanziale sovrapposizione del testo (cfr. I,1 Coro Vale Magister maxime; I,1 Coro Vale Magister maxime). Il [1803] è sostanzialmente una riscrittura di [1784]. Libretto identificato per l'identità di nome e didascalia di un solo personaggio: "mastro Giorgio, chirurgo e maestro de pazzi".
	Tipo di relazione
	Rifacimento

Caso:6	Opere individuate
	<i>Stratonica</i> , ?-Salvi, Firenze, 1707 (http://corago.unibo.it/resource/BRANI/APC0003750)
	Rappresentazione: <i>Stratonica</i> , ?-Salvi, Firenze, 1707 (http://corago.unibo.it/resource/SPETTACOLI/SSC0001158)
	Rappresentazione: <i>Stratonica</i> , ?-Salvi, Livorno, 1717 (http://corago.unibo.it/resource/SPETTACOLI/SST0041523)
	Criteri utilizzati
	Personaggi: Dorisbe, vecchia nutrice d'Arsinoe; Lesbo, servo di Tigrane.
	Analisi critica

	<p>In [FI1707] i due personaggi buffi compaiono soltanto in scene specifiche all'interno degli atti: I,9; I,15; II,8; III,4.</p> <p>In [FI1717] Lesbo scompare dall'elenco dei personaggi e Dorisbe è sostituita da Zafira, damigella d'Arsinoe. In calce all'elenco dei personaggi compare l'indicazione "Intermezzi", ma senza nomi di personaggi; il testo di Dorisbe della I,9 di [FI1707], identico, è affidato a Zafira; le scene dei due buffi di [FI1707] e la menzione nell'ultima scena sono state cassate.</p> <p><i>Il libretto Intermezzi comici musicali</i> (http://corago.unibo.it/resource/DOCUMENTI/DRT0024341), [FI1717] per lo spettacolo di [LI1717] contiene l'intermezzo Lesbo e Zafira; il testo è diverso dalle scene buffe di [FI1707], ma è senz'altro l'intermezzo per lo l'opera <i>Stratonica</i> [LI1717], soltanto che questa volta il libretto è stato stampato a parte.</p>
	Tipo di relazione
	Rielaborazione

Caso:7	Opere individuate
	<p><i>Giove in Argo</i>, Lotti A.-Lucchini A. M., Dresda, 1717 (http://corago.unibo.it/resource/BRANI/0000302551) Libretto: <i>Giove in Argo</i>, Dresda, 1717 (http://corago.unibo.it/resource/DOCUMENTI/DRT0021821)</p> <p><i>Cleonice</i>, Ristori G.A.-Costantini A., Moritzburg, 1718 (http://corago.unibo.it/resource/BRANI/0000362473) Libretto1: [<i>La Cleonice</i>], Dresda, [1718?] (http://corago.unibo.it/resource/DOCUMENTI/0001140146) Libretto2: <i>La Cleonice</i>, Verona, 1723 (http://corago.unibo.it/resource/DOCUMENTI/DRT0010551)</p>
	Criteri utilizzati
	Personaggi: Cleone: pastore; Erasto: pastore riconosciuto per Osiri re d'Egitto.
	Analisi critica
	<p>Le porzioni di testo di [DR1718] non sono della <i>Cleonice</i>, ma del <i>Giove in Argo</i>: porzioni degli atti I.6, II.6 e III relativi alle parti di Cleone ed Erasto. Vi compaiono delle note manoscritte di richiamo ("Vedi A" ... B, C, D), probabilmente aggiunte dallo stampatore, che vanno intese come riferimenti per l'inserzione, nel libretto difettoso, del testo delle pagine stampate a parte.</p> <p>Su questa base documentaria e dall'osservazione dell'esemplare si può argomentare: che il libretto <i>Giove in Argo</i> 1718 è difettoso; che lo stampatore si sia accorto dell'errore di stampa (omessa stampa di una parte del testo) soltanto dopo aver fascicolato il libretto; per porre rimedio all'errore ha provveduto alla stampa di un nuovo fascicolo contenente le parti di testo omesse; nel libretto difettoso ha poi aggiunto le note manoscritte A,B,... per indicare al lettore un riferimento per l'inserzione del testo pubblicato nel fascicolo a parte.</p>

	Ne deriva che l'edizione catalogata dalla biblioteca di Dresda come <i>Cleonice</i> quindi non esiste, è un errore di catalogazione: il primo libretto conosciuto della <i>Cleonice</i> è quindi [VR1723].
	Tipo di relazione
	Ricostruzione unità bibliografica mediante attestazione di due fascicoli separati ad una unica edizione.

Caso:8	Opere individuate
	<p><i>La fede riconosciuta</i>, Pollarolo-[Pasqualigo], Vicenza, 1707 (http://corago.unibo.it/resource/BRANI/APC0003372)</p> <p><i>La fede riconosciuta</i>, Scarlatti-B. Marcello, Napoli, 1710 (http://corago.unibo.it/resource/BRANI/0000302379)</p> <p><i>Dorinda</i>, G.B. Pascetti, B. Galuppi-Lalli, Venezia, 1729 (http://corago.unibo.it/resource/BRANI/APC0004453)</p> <p><i>Dorinda</i>, F. Vitali-Tagliazucchi, Sassuolo, 1740 (http://corago.unibo.it/resource/BRANI/0001405476)</p>
	Criteri utilizzati ed indizi ricavati
	Personaggio: Dorinda, ninfa di Delo creduta Fileno, amante di Silvio.
	Analisi critica
	<p><i>La fede riconosciuta</i> [NA1710] è erroneamente attribuita a B. Marcello in (Groppo, 1745, p. 128) essendo una ripresa di [VI1707]: compaiono infatti varianti poco rilevanti nei recitativi e vengono riprese alcune arie.</p> <p>La <i>Dorinda</i> [VE1729] è attribuita dal Grove al Lalli (Hill, 2001); in (Bonlini, 1730, p. 213) è riportato per [VE1729] che "Questo dramma, che fu riformato, è quello che col titolo della Fede Riconosciuta fu recitato a Vicenza, correndo ivi la Fiera di Maggio dell'anno 1707". Da riconsiderare quindi l'attribuzione al Lalli attestata nel Grove.</p> <p>La <i>Dorinda</i> [VE1740] risulta a sua volta essere una rielaborazione di [VE1729] essendo ripreso con alcune varianti l'atto I, rifatti gli atti II e III. Va quindi aggiunto il riferimento al Pasqualigo come librettista.</p>
	Tipo di relazione
	[NA1710] ripresa di [VI1707]; [VE1740] rielaborazione di [VE1729].

I casi riportati illustrano le diverse casistiche in cui il metodo proposto permette di individuare i diversi tipi di legami tra opere. Gli ultimi tre casi evidenziano come lo stesso metodo risulti efficace anche in situazioni in cui le relazioni sussistano tra contesti testuali diversi, in vicende editoriali travagliate o in presenza di tradizioni incomplete frutto di attribuzioni errate.

Il caso 6 evidenzia come il metodo funziona anche se il contesto testuale in cui i personaggi compaiono è diverso. I due personaggi buffi Dorisbe e Lesbo, perfettamente integrati nell'azione drammatica di una opera seria come *Stratonica* nel 1707, vengono invece estratti e trasposti nell'intermezzo *Lesbo e Zafira* del 1717. La progressiva espunzione dei personaggi comici dal dramma serio comincia all'inizio del secolo per compiersi nei primi due decenni dando vita ad un genere autonomo (Troy e Weiss, 2001).

Il caso 7 mostra come il metodo sia efficace anche nell'individuare contenuti resi spuri da un errore di stampa che ha a sua volta indotto un errore di catalogazione generando così una prima attestazione inesistente: alla luce di questo non si ha un'evidenza della *Cleonice* prima del 1723. Non è certo possibile risalire alla ragione dell'attestazione del libretto spurio, per altro chiaramente annotata a matita nel relativo frontespizio dell'esemplare di Dresda: unica ipotesi è l'assonanza tra il titolo dell'opera *Cleonice* e il personaggio Cleone.

Il caso 8 illustra invece l'efficacia nel colmare le lacune nella tradizione del testo fino ad oggi nota correggendo attribuzioni erranee. Il risultato è la ricostruzione di una catena che lega il libretto de *La fede riconosciuta* del Pasqualigo [1707] alla *Dorinda* rappresentata a Sassuolo nel 1740.

Infine, il caso 4 evidenzia come il metodo permetta di rintracciare la relazione tra due libretti dal titolo completamente diverso lontani l'uno da l'altro centoventidue anni in assenza di testimoni intermedi.

3.2.4 Conclusioni preliminari ed evoluzioni future

Nel capitolo sono state presentate le tre dimensioni attraverso cui è divenuto possibile accedere al patrimonio informativo dell'archivio. Le prime due, spazio e tempo, sono dimensioni proprie del dato storico; la terza, il personaggio, è stata disegnata come strumento di accesso ai testi del teatro d'opera. I benefici derivanti dall'unificazione dei riferimenti temporali e geografici dei contenuti dell'archivio verranno evidenziati nel capitolo 5 a valle dell'illustrazione delle funzionalità di consultazioni che la sfruttano.

I risultati derivanti dall'impiego della dimensione personaggio, per quanto preliminari, evidenziano invece una significativa efficacia nell'individuare indizi di legami tra testi. L'attività di popolamento del modello descrittivo si è limitata al primo livello, quello che ha come riferimento la 'figura drammatica'. L'estrazione di attributi qualificati in modalità parzialmente supervisionata, che viene descritta nel capitolo 4, è il passaggio preliminare per procedere al popolamento del livello successivo relativo agli 'archetipi'. In questa prospettiva dovranno essere esplorate le metodologie necessarie ad individuare criteri di aggregazione. A questo scopo si dovranno esplorare metodi di *clustering* sia come supervisionati che non (Rojas, 1996). Tra i secondi, sono state effettuate alcune sperimentazioni preliminari utilizzando algoritmi di *clustering* agglomerativo basate sul metodo K-means che hanno evidenziato significative potenzialità (He et al., 2006).

4 Estensione del patrimonio informativo

Nel capitolo 2 è stato illustrato il processo di trasposizione dei contenuti dell'archivio verso un modello concettuale la cui semantica viene formalmente esplicitata con l'obiettivo di rispondere a nuovi requisiti in termini di accessibilità dei dati. Nel capitolo 3 sono state approfondite le potenzialità derivanti dall'impiego di dimensioni di analisi in grado di valorizzare il patrimonio informativo già presente nell'archivio.

In questo capitolo vengono invece presentate le metodologie ed i relativi strumenti utilizzati per analizzare i contenuti testuali dell'archivio con l'obiettivo di estenderne il patrimonio informativo. Questa estensione può essere perseguita sia con l'introduzione di nuovi contenuti provenienti da fonti esterne che con l'estrazione di informazione già disponibile in forma implicita (ovvero non ancora formalizzata nel modello descrittivo).

Ad esempio, rientrano in questa seconda tipologia le informazioni della collocazione geografica delle ambientazioni relative alle opere o le relazioni tra i personaggi descritte nelle didascalie. In entrambi i casi né il modello relazionale originario, né la sua trasposizione nel modello basato su ontologie prevedevano l'esplicitazione di queste informazioni.

Sono state sperimentate due diverse metodologie: l'analisi automatica dei testi presenti nell'archivio e l'allineamento degli stessi con fonti esterne in cui reperire informazioni correlate alle risorse già censite.

Il primo metodo consiste nell'estrarre informazione qualificata mediante l'analisi automatica dei contenuti espressi in linguaggio naturale (*Natural Language Processing* - NLP). Partendo da un'analisi morfosintattica dei contenuti testuali effettuata con strumenti NLP sono state sperimentate delle euristiche di estrazione che hanno consentito di arricchire le descrizioni di entità come i personaggi e le opere.

La seconda metodologia sfrutta l'allineamento tra basi di conoscenza LOD con l'obiettivo di derivare dalle risorse esterne nuovi attributi da associare alle risorse locali. Questa metodologia è stata impiegata per l'acquisizione delle coordinate geografiche delle località.

I casi di studio utilizzati per sperimentare l'efficacia delle due metodologie e dei relativi strumenti sono stati pensati per ottenere l'estensione del perimetro informativo necessaria per evolvere le funzionalità di navigazione e di interrogazione dell'archivio descritte nel capitolo 5.

4.1 Stato dell'arte delle metodologie di analisi automatica dei contenuti

L'impiego di strumenti informatici per l'estrazione di dati strutturati a partire da informazioni presenti in testi espressi in linguaggio naturale ha ormai una storia più che trentennale (Martin e Jurafsky, 2009, p. 796). L'insieme di tecniche che svolgono questo tipo di analisi dei testi costituiscono il campo di indagine dell'*information extraction* (IE). Tra i temi affrontati in questo ambito sono di nostro interesse: l'individuazione di entità note all'interno di un testo (*Named Entities Recognition* - NER), l'estrazione di relazioni tra queste entità e una loro classificazione (*Relation Detection and Classification* - RDC) e l'estrazione di caratteristiche riferite alle entità individuate (*Feature Extraction* - FE).

Queste tecniche di analisi hanno l'obiettivo comune di riuscire ad estrarre informazione rappresentabile in forma strutturata. Dovendo trattare testi complessi, le attuali applicazioni di IE sono costruite attraverso una pila (o *stack*) di componenti che effettuano in successione una sequenza di analisi che, partendo dalla superficie lessicale, arrivano all'annotazione morfosintattica ed alla lemmatizzazione dei contenuti. Quest'ultimo livello di annotazione del testo consente la costruzione di processi di estrazione basati

sull'aderenza dei contenuti a specifici schemi sintattici di riferimento (*syntactic paths*) e termini in essi contenuti.

Questa modalità di estrazione dell'informazione basata su *pattern* morfosintattici rientra nell'ambito delle tecniche cosiddette a bassa supervisione (*Lightly Supervised Approaches* - LSA) (Martin e Jurafsky, 2009, p. 772). Le LSA sono applicazioni di estrazione dell'informazione che richiedono una ridotta supervisione dell'attore umano in fase di addestramento del sistema. La supervisione è infatti limitata alla sola definizione delle regole di estrazione definite in base allo specifico ambito di applicazione.

4.1.1 Analisi NLP e tagging morfosintattico

Le tecniche di analisi automatica del testo risalgono ai primi anni Sessanta¹⁰⁸. In particolare, a partire dagli anni Ottanta, il problema dell'individuazione delle componenti della frase (Part of Speech Tagging - PoS) è stato affrontato con crescente successo attraverso l'applicazione di modelli statistici addestrati su corpora annotati. Impiegate dapprima all'analisi delle singole parti del discorso, sono state applicate poi all'analisi della struttura sintattica della frase rappresentata formalmente sotto forma di dipendenze funzionali¹⁰⁹ (De Marneffe e Manning, 2008). A questo livello di annotazione la frase viene rappresentata come una sequenza di relazioni di dipendenza tra i termini che la compongono. La struttura che ne deriva è facilmente approssimabile ad un grafo formato da triple aventi per nodi i termini e come archi la relazione di dipendenza. I diversi tipi di relazione sintattica vengono raccolti in dizionari specifici per lingua. Con l'introduzione delle *Universal Dependencies* (UD) si cerca di superare questo vincolo introducendo una tassonomia di dipendenze multilingua (De Marneffe et al., 2014). L'introduzione delle UD consente di estendere l'annotazione anche all'italiano per cui è stato effettuato l'adeguamento delle librerie di annotazione morfosintattica Stanford Core NLP (Aprosio e Moretti, 2016)¹¹⁰. Questo passaggio ci ha permesso di sperimentare euristiche di estrazione di informazione basate su *pattern* sintattico-lessicali applicate ai contenuti testuali presenti in *Corago*.

4.1.2 Estrazione dell'informazione non supervisionata basata su *pattern*

L'efficacia degli algoritmi per l'estrazione automatica di informazioni viene misurata in base alla capacità di reperire nuova informazione (*recall*) e alla precisione ed aderenza al dominio di interesse dell'informazione estratta (*precision*). Nei sistemi basati sull'apprendimento di modelli statistici a partire da corpora annotati (approccio supervisionato da attori umani) la precisione è direttamente legata alla dimensione ed alle caratteristiche del corpus di riferimento. Al tempo stesso la *recall* dipende dalla varietà dei campioni utilizzati per l'addestramento in relazione all'ampiezza dell'ambito di interesse.

Nel nostro caso il tipo di caratteristiche (*features*) e di relazioni che si vogliono individuare non sono necessariamente note a priori. Non è quindi possibile utilizzare criteri di estrazione basati su esempi attestati in un corpus. I sistemi per l'annotazione delle dipendenze sintattiche offrono una possibile alternativa (Mintz et al., 2009). Analizzando le caratteristiche lessicali del testo e la sua articolazione sintattica è possibile individuare strutture ricorrenti nel discorso (*patterns*) in base alle quali definire le euristiche di estrazione. Accoppiando l'estrazione dei *pattern* a specifici lemmi utilizzati come criteri di restrizione è possibile

¹⁰⁸ Si rinvia alla approfondita trattazione in (Martin and Jurafsky, 2009) che nel capitolo introduttivo disegnano l'evoluzione dell'analisi automatica del linguaggio naturale dal 1950 agli anni Duemila.

¹⁰⁹ La proposta di una grammatica libera dal contesto (Context Free Grammar - CFG) in cui la frase viene rappresentata come una concatenazione di dipendenze sintattiche tra le parole che la compongono viene fatta risalire ai primi lavori di Noam Chomsky negli anni Cinquanta (Martin and Jurafsky, 2009, p. 457).

¹¹⁰ Per approfondimenti sulla localizzazione dello stack Stanford CoreNLP per l'italiano si rinvia al sito web relativo al progetto Tint promosso dalla Fondazione Bruno Kessler: <http://tint.fbk.eu/> (visionato 24/03/2019).

raggiungere significativi livelli di *recall* mantenendo alta la *precision* pur avendo adottato un approccio non supervisionato (Moro et al., 2013).

Nel nostro caso è stata sperimentata una combinazione dei diversi strumenti fin qui descritti: l'annotazione delle dipendenze morfosintattiche; le euristiche di estrazione basate su *pattern* e l'estrazione vincolata alla semantica dei termini contenuti. Nell'insieme l'approccio seguito estrae informazione qualificata sulla base di criteri semantici applicati al grafo sintattico del testo sottoposto ad analisi.

4.2 Estrazione dell'informazione dai contenuti esistenti

L'estrazione di informazione già presente in forma non strutturata nell'archivio ha l'obiettivo di far emergere informazioni presenti nei contenuti testuali, ovvero in una forma non esplicitata dal modello concettuale di riferimento, e quindi non utilizzabile direttamente come chiave di ricerca o dimensione di analisi.

Per rendere esplicita l'informazione presente nei testi è necessario elaborare una procedura di estrazione strutturata su più fasi (ovvero una *pipeline* di estrazione). In primo luogo, si dovrà processare la superficie lessicale del testo per individuarne le parti del discorso e le relative relazioni sintattiche. Ottenuti questi due livelli di annotazione del testo, si può procedere con l'individuazione delle porzioni il cui contenuto è rilevante rispetto al tipo di informazione ricercata. Individuate le unità lessicali di interesse, si dovrà ricavarne una rappresentazione dei contenuti informativi strutturata in una forma coerente al modello concettuale della base di conoscenza.

Data la necessità di combinare strumenti disegnati per un impiego di tipo generalista ad euristiche di estrazione disegnatte per estrarre contenuti specifici, è stata realizzata una *pipeline* dedicata a ciascuno dei due casi di studio sperimentati.

4.2.1 I casi di studio: descrizioni di personaggi ed ambientazioni

Nel capitolo 3 sono state presentate tre dimensioni di analisi particolarmente efficaci come chiavi di accesso al dominio. Due di queste sono ricavabili da informazioni testuali: le didascalie dei personaggi e il testo che descrive le ambientazioni delle opere.

Nel paragrafo 3.2.2.1 è stato definito il concetto di 'attributo del personaggio' inteso come una porzione di testo che esprime una specifica caratteristica del personaggio. Il contenuto informativo di una istanza della classe `corago:C17_Character_Feature` dovrà essere un termine, sintagma o proposizione estratto dalla didascalia del personaggio. Ad esempio, Cherinto¹¹¹ viene descritto come "figlio di Demofoonte, amante di Creusa". Le due *features* del personaggio che si possono ricavare sono: Cherinto è il "figlio di Demofoonte" e l'"amante di Creusa". Un successivo livello di analisi dovrà quindi individuare le relazioni con i due personaggi riferiti e formalizzarne la rappresentazione attestandola sull'ontologia di dominio. Con quest'ultimo passaggio si passa dall'informazione '*human readable*' della didascalia alla sua espressione formalmente codificata e quindi '*machine readable*'.

Nel caso delle ambientazioni delle azioni drammatiche delle opere, quello che ci interessa individuare sono riferimenti a località in base a cui poter collocare geograficamente l'azione drammatica rappresentata sulla scena. La nota relativa all'ambientazione de *Il duca d'Alba*¹¹², opera composta da Gaetano Donizetti su libretto di Matteo Salvi e rappresentata a Roma nel 1882, ci informa che: "l'azione si svolge a Brusselle e nel porto di Anversa, sulla riva della Schelda. Epoca: 1573". In questo caso la procedura di estrazione dovrà essere in grado di individuare "Brusselle", "Anversa" e "la Schelda" come toponimi. Un primo livello di annotazione

¹¹¹ Riferimento URI del personaggio: <http://corago.unibo.it/resource/PERSONAG/0000050806>.

¹¹² Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/Z000002076>.

deve identificarli come sostantivi, un secondo livello come potenziali toponimi e, infine, un terzo livello dovrà individuare la località geografica del toponimo per ricavare le coordinate geografiche. Il risultato sarà una rappresentazione strutturata della collocazione geografica dell'ambientazione dell'opera.

4.2.2 Analisi NLP dei contenuti testuali

Il primo passaggio per la realizzazione di una *pipeline* di estrazione automatica di contenuti espressi in linguaggio naturale è l'identificazione delle caratteristiche morfosintattiche del testo. Questa operazione viene realizzata attraverso l'annotazione linguistica del testo che consente di codificare l'informazione linguistica relativa alle diverse componenti del testo. Questo è il passaggio chiave che permette di «rendere esplicita, interpretabile ed esplorabile la struttura linguistica implicita nel testo» (Lenci et al., 2005, p. 211).

L'annotazione del testo copre aspetti diversi dell'informazione linguistica. La forma basilare di annotazione che costituisce il presupposto per i livelli successivi è quella morfo-sintattica che assegna ad ogni parola (o *token*) la rispettiva categoria grammaticale, ovvero la parte del discorso (Part of Speech - PoS) o categoria lessicale (*ibidem*, p. 213). A questa si affianca la lemmatizzazione, ovvero l'identificazione dell'esponente lessicale (o lemma) del singolo termine. La lemmatizzazione permette di operare sul lessico astraendo dalle variazioni di tipo morfologico naturalmente presenti nel testo: ad esempio è possibile trattare le forme varianti derivate di “figlia”, come “figlie” o “figliola”, in modo analogo.

Il secondo livello di annotazione linguistica è rappresentato dall'annotazione sintattica che codifica le informazioni relative alla struttura sintattica della frase del periodo analizzato. A questo livello diventa determinante il modello teorico cui fa riferimento il sistema di annotazione. Nel nostro caso si è scelto di adottare l'approccio basato sulla rappresentazione delle dipendenze per la sua capacità di rappresentare la struttura sintattica in termini di relazioni binarie facilmente esprimibili in forma di triple. Questo ci permetterà di trattare la struttura sintattica sotto forma di grafo e quindi di implementare i *pattern* di estrazione come interrogazioni espresse in linguaggio SPARQL.

Riprendendo il caso di Cherinto “figlio di Demofoonte, amante di Creusa” il livello di annotazione morfosintattica della frase è riportato in tabella 4.1.

Tabella 4.1 esempio di annotazione morfo-sintattica di una didascalia di personaggio

Parola	Lemma	PoS ¹¹³
Cherinto	Cherinto	PROPN (nome proprio)
figlio	figlio	NOUN (sostantivo)
di	di	ADP (preposizione)
Demofoonte	Demofoonte	PROPN (nome proprio)
,	[PUNCT]	PUNCT (segno di interpunzione)
amante	amante	NOUN (sostantivo)
di	di	ADP (preposizione)
Creusa	Creusa	PROPN (nome proprio)

¹¹³ Viene qui riportata la forma trasposta nel *tagset* formato CoNLL-2007, tra parentesi il significato del *tag*.

A questo livello di annotazione non è ancora possibile individuare il nesso tra il soggetto e i predicati nominali che gli sono riferiti. È necessario quindi procedere a livello di annotazione sintattica per essere in grado di associare il soggetto e i termini che lo qualificano.

Parola	Dipendenza	Parola
Cherinto	< http://universaldependencies.org/u/dep#amod >	figlio
di	< http://universaldependencies.org/u/dep#case >	Demofoonte
Demofoonte	< http://universaldependencies.org/u/dep#nmod >	figlio
,	< http://universaldependencies.org/u/dep#punct >	figlio
amante	< http://universaldependencies.org/u/dep#appos >	figlio
di	< http://universaldependencies.org/u/dep#case >	Creusa
Creusa	< http://universaldependencies.org/u/dep#nmod >	amante

Tabella 4.1: Dipendenze tra termini espresse nella codifica UD

La figura 4.1 illustra la concatenazione delle relazioni binarie espresse in base al dizionario delle UD.

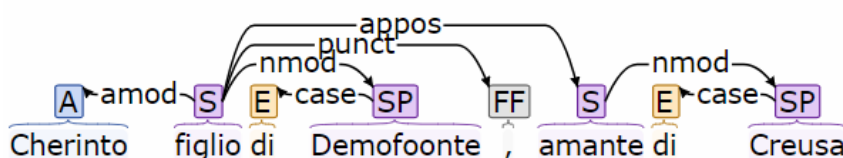


Figura 4-1: Esempio di annotazione sintattica UD

Partendo dalla testa della frase, Cherinto, la prima relazione individuata è tra Cherinto e il sostantivo “figlio” identificato come modificatore aggettivale¹¹⁴ (“amod” è l’etichetta utilizzata per l’annotazione) del primo termine. A sua volta “figlio” è associato a “amante” dalla relazione di apposizione (etichetta “appos”), ovvero di modificatore coordinato a “figlio” che ne estende la definizione. Ciascuno dei due modificatori “figlio” e “amante” hanno a loro volta dei modificatori: “Demofoonte” specifica di chi è “figlio” Cherinto e “Creusa” di chi è amante. Se si scorre la gerarchia delle dipendenze possiamo attribuire al personaggio Cherinto due attributi considerando i modificatori aggettivali, ovvero che è contemporaneamente “figlio di Demofoonte” e “amante di Creusa”.

¹¹⁴ La guida alle relazioni sintattiche definite nelle UC versione 2.x è mantenuta online all’indirizzo: <https://universaldependencies.org/u/dep/> (verificato il 24/03/2019).

Nell'ipotesi formulata, l'uso della struttura morfosintattica dei periodi contenuti nella didascalia (rappresentata sotto forma di grafo di dipendenze funzionali tra i termini che la compongono) permette di segmentare i contenuti della descrizione del personaggio in sotto-unità informative. Ciascuna di queste può essere assunta come un attributo del personaggio e rappresentata come istanza della classe `corago:C17_Character_Feature`¹¹⁵. Oppure potrà essere ulteriormente segmentata fino al raggiungimento della massima granularità corrispondente al singolo termine.

L'individuazione delle porzioni di testo da considerare come attributi viene effettuata in base a *pattern* sintattici espressi attraverso le UD. Ad esempio, il grafo sintattico dell'esempio relativo a Cherinto, suggerisce che la struttura articolata in testa della frase, modificatore aggettivale, nome e relativi modificatori, è in grado di estrarre *features* relative al personaggio in posizione di testa del predicato. Si dovrà quindi individuare un formalismo in grado di esprimere la struttura del sintagma tipo oggetto della ricerca, ovvero il *pattern* ricercato. Il processo di estrazione delle porzioni di testo corrispondenti dovrà limitarsi ad applicare il *pattern* espresso in questa forma.

Per verificare l'ipotesi iniziale è necessario individuare un insieme di *pattern* di estrazione di *features* specifiche e valutare il significato del contenuto estratto. A questo scopo si deve introdurre un ulteriore livello di analisi che consideri la semantica del lessico contenuto nei sintagmi estratti in base ai *pattern*. La *pipeline* di estrazione dovrà perciò essere articolata su più livelli di annotazione ed eseguita in più fasi.

Il primo passaggio consiste nel produrre l'annotazione morfosintattica del testo di partenza. Il secondo consisterà nell'estrazione di sezioni le cui dipendenze corrispondono alla struttura del sintagma definita dal *pattern* di estrazione. Ottenuti i sintagmi si può procedere a valutare la coerenza semantica dei contenuti utilizzando il criterio lessicale. Ad esempio, nel caso del personaggio Cherinto, si dovrà verificare che i sostantivi che fungono da modificatori aggettivali rientrino in una lista controllata di termini. In questo modo, se si elencano lemmi che esprimono una classe coerente di contenuti, ad esempio quelli di relazione parentale di primo grado (come "figlio", "figlia", "madre", "padre"), si potranno identificare i sintagmi che descrivono la genealogia del personaggio.

Una procedura di estrazione così strutturata permette di indagare i contenuti testuali delle descrizioni identificandone porzioni che corrispondono sia a specifiche caratteristiche sintattiche sia a criteri relativi al contenuto. Per ottenere questo risultato è però necessario realizzare un modello concettuale che consenta di operare contemporaneamente sui tre livelli di annotazione: morfologico (PoS e lemmatizzazione), sintattico (annotazione delle dipendenze con le UD) e lessicale. Diventa così possibile specificare criteri di estrazione che individuano le caratteristiche dei personaggi in base alla semantica del lessico contenuto.

4.2.3 Annotazione morfosintattica: UD in RDF

La necessità di definire *pattern* di estrazione che prevedono vincoli organizzati su livelli gerarchici di annotazione richiede che i *pattern* vengano espressi in forma strutturata attraverso un modello in grado di rappresentare il diverso contenuto informativo di ciascuno di essi. Si è deciso quindi di rappresentare sia il testo che i livelli di annotazione utilizzando un grafo RDF¹¹⁶.

¹¹⁵ Si rinvia al paragrafo 3.2.2.2 per la definizione della classe.

¹¹⁶ Il componente software sviluppato per realizzare l'annotazione morfosintattica in formato RDF è descritto nel paragrafo 2.2 dell'Appendice C.

Il risultato è un *dataset* in cui sono rappresentati i personaggi, le rispettive didascalie ed il risultato della procedura di annotazione NLP. A questo scopo si è realizzata una versione in OWL del dizionario delle Dipendenze Universali arricchito delle entità di dominio e delle relazioni necessarie a completare il modello.

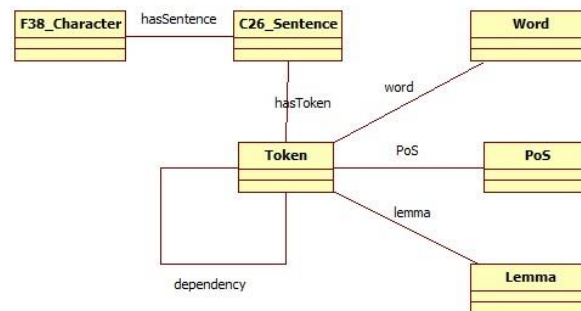


Diagramma 4.1: Modello annotazioni morfosintattiche

In base al modello riportato nel diagramma 4.2, ciascuna didascalia viene rappresentata come insieme ordinato di *token*. I tre livelli di annotazione di ciascuno di essi sono rappresentati attraverso specifiche proprietà: l'estrazione della superficie lessicale (ud:word), il tag PoS (ud:POS), il lemma (ud:lemma). I *token* sono tra di loro correlati dalle relative dipendenze (ud:dependency).

Il grafo ottenuto con la trasposizione dei livelli di annotazione in formato RDF è interrogabile col linguaggio SPARQL. L'espressività del linguaggio permette di specificare sia composizione sia la natura delle dipendenze tra i *token*. Permette inoltre di imporre vincoli sul contenuto dei nodi a contenuto letterale. In questo modo è possibile costruire criteri di interrogazione del grafo che specifichino contemporaneamente sia la struttura sintattica della frase (sequenze di relazioni di dipendenza) sia i contenuti lessicali (valori dei nodi *lemma*).

In questo modo si raggiunge l'espressività necessaria a costruire una regola di estrazione che, ad esempio, identifichi tutti i personaggi che hanno una relazione di parentela con un altro personaggio dell'opera.

4.2.4 Estrazione basata su regole

La possibilità di estrarre informazioni dalle didascalie dei personaggi in base a regole che analizzano contemporaneamente la struttura sintattica (mediante *pattern* morfosintattici) e la semantica dei contenuti (attraverso *pattern* lessicali) è stata verificata sperimentalmente.

Il caso di studio scelto è rappresentato dalle relazioni che intercorrono tra i personaggi di una stessa opera attestate nelle didascalie. Dagli esempi riportati in 4.2.2 si evince come la relazione tra due personaggi sia articolata a livello sintattico e qualificata dai termini attraverso cui è espressa. Il modello concettuale dell'annotazione presentato in 4.2.3 permette di coniugare i due piani. Le regole per l'estrazione di relazioni saranno quindi composte da:

1. un *pattern* sintattico che definisca la struttura dei sintagmi che stabiliscono una relazione tra due personaggi;
2. un dizionario che restringa i contenuti del sintagma al dominio di interesse.

Il *pattern* ha il compito di identificare i sintagmi che mettono in relazione il personaggio 'soggetto' della relazione con il personaggio 'oggetto'. Nei nostri esempi, Cherinto (soggetto) "è figlio di" Demofonte (oggetto). Il dizionario, inteso come elenco di lemmi che definiscono il tipo di relazione, permette di discriminare la qualifica di "figlio" da quella di "amante". La prima può essere classificata come relazione di parentela, la seconda come relazione interpersonale.

Nel nostro caso deve essere previsto un terzo criterio di estrazione che consenta di identificare il personaggio in posizione di 'oggetto' nella relazione come uno dei personaggi della stessa opera. In questo caso si dovrà verificare che il lemma corrispondente coincida con un nome presente nell'elenco dei personaggi dell'opera.

La costruzione di regole viene fatta utilizzando due strumenti di base: un insieme di *pattern* che rappresentano le diverse varianti di dipendenze sintattiche con le quali una relazione può essere espressa e un insieme di lemmi che qualificano la tipologia di relazione oggetto della ricerca.

In termini generali, ci si aspetta che con l'estensione del primo si aumenterà la *recall* complessiva della regola, ovvero la capacità di selezionare un numero più vicino a tutti i sintagmi che specificano una relazione utile. Restringendo il numero di elementi del secondo insieme si aumenterà la *precision*, ovvero il controllo sulla coerenza semantica della relazione estratta rispetto alla tipologia di interesse.

La definizione della struttura dei *pattern* può essere fatta sia in modo deduttivo che empirico-induttivo. Nel primo caso si ipotizzano le catene di dipendenze che si reputano possano contenere relazioni utili, nel secondo si analizza un campione di relazioni note e se ne deriva la struttura. Nel nostro caso si è optato per il secondo approccio applicando delle generalizzazioni dove opportuno¹¹⁷.

¹¹⁷ Un esempio di applicazione dello stesso approccio per la definizione di regole di estrazione basate su una combinazione di lessico e sintassi è reperibile in (Poria et al., 2014).

```

#Interrogazione SPARQL per l'estrazione della relazione padre/figlio tra personaggi
CONSTRUCT
{
  #BLOCCO 0: costruzione della tripla che specifica la relazione tra i personaggi
  ?characterSbjURI <http://www.co-ode.org/roberts/family-tree.owl#isSonOf> ?characterObjURI.
}
WHERE
{
  #BLOCCO 1: identificazione del personaggio in posizione di 'soggetto'
  BIND(<http://corago.unibo.it/resource/PERSONAG/0000050806> as ?characterSbjURI)

  #BLOCCO 2: estrazione didascalia del personaggio
  ?characterSbjURI <http://corago.unibo.it/sm/hasSentence> ?sent.
  ?sent rdfs:label ?lblSent.
  ?characterSbjURI ^<http://corago.unibo.it/sm/CNLP1_refers_to_character> ?1.
  ?1 <http://universaldependencies.org/u/dep#word> ?characterName.

  #BLOCCO 3: estrazione TOKEN cui applicare il pattern
  ?sent <http://universaldependencies.org/u/dep#hasToken> ?1.
  ?sent <http://universaldependencies.org/u/dep#hasToken> ?2.
  ?sent <http://universaldependencies.org/u/dep#hasToken> ?3.

  #BLOCCO 4: definizione del pattern sintattico -> inverse(ud:amod)/ud:nmod/ud:case con PoS(ud:amod)=NOUN
  ?1 ^<http://universaldependencies.org/u/dep#amod> ?2.
  ?2 <http://universaldependencies.org/u/dep#nmod> ?3.
  ?3 <http://universaldependencies.org/u/dep#case> ?3c.
  ?2 <http://universaldependencies.org/u/dep#POS> "NOUN".

  #BLOCCO 5: estrazione lemmi corrispondenti
  ?1 <http://universaldependencies.org/u/dep#lemma> ?1l.
  ?2 <http://universaldependencies.org/u/dep#lemma> ?1l2.
  ?3 <http://universaldependencies.org/u/dep#lemma> ?1l3.
  ?3c <http://universaldependencies.org/u/dep#lemma> ?1l3c.

  #BLOCCO 6: imposizione vincolo sul lessico
  FILTER(?1l2 IN ('figlio'))

  #BLOCCO 7: estrazione personaggio in posizione di 'oggetto'
  SERVICE <http://coragolod2.ing.unibo.it:8080/blazegraph/namespace/corago_ft_beta/sparql>
  {
    ?characterObjURI a <http://erlangen-crm.org/efrbroo/F38_Character>.
    ?work <http://erlangen-crm.org/current/P129_is_about> ??characterSbjURI.
    ?work <http://erlangen-crm.org/current/P129_is_about> ?characterObjURI.
    ?characterObjURI rdfs:label ?1l3.
  }
}

```

Listato 4.1: Interrogazione SPARQL per l'estrazione della relazione padre/figlio

Per la definizione degli elenchi dei lemmi ‘qualificanti’ il tipo di relazione si è ricorso all’utilizzo di dizionari¹¹⁸. Navigando la gerarchia di iponimi di un termine qualificante, si è ottenuto un elenco di lemmi il cui significato rappresentasse efficacemente la tipologia di relazione.

Il listato 4.1 riporta un esempio di interrogazione SPARQL relativo al caso del personaggio Cherinto che estrae dal grafo annotato la relazione “figlio di Demofonte”¹¹⁹.

¹¹⁸ Come dizionario per i lemmi è stato utilizzato il *synset Babelnet* (Navigli e Ponzetto, 2012).

¹¹⁹ In questo caso rappresentata formalmente dalla proprietà *famtree:isSonOf*.

Dal punto di vista implementativo la regola di estrazione viene definita come una istruzione di tipo CONSTRUCT del linguaggio SPARQL. Questo tipo di istruzione restituisce un insieme di triple che asseriscono uno o più fatti (o assiomi) ricavati dalla base di conoscenza (blocco 0) in base ai criteri definiti nella corrispettiva clausola di WHERE. Questa è a sua volta articolata in più blocchi con funzioni specifiche.

Nell'esempio riportato, al blocco 1 viene assegnato il valore dell'URI del personaggio di cui si sta analizzando la didascalia. Nel blocco 2 vengono estratte le variabili relative ai contenuti della didascalia. Nel blocco 3 vengono identificati i termini (sotto forma di *token*) che partecipano alla struttura del sintagma definita come *pattern* di relazioni UD tra di essi nel blocco 4. Il successivo blocco 5 estrae i lemmi corrispondenti ai *token* e nel blocco 6 viene imposto il filtro sul contenuto lessicale del *token* più significativo: in questo caso il sostantivo modificatore aggettivale. Nel blocco 7 viene imposto il vincolo che il *token* modificatore nominale corrisponda al nome di un personaggio della stessa opera.

In termini più generali le clausole di estrazione devono prevedere tre sezioni fondamentali, qui rappresentate dai blocchi 4, 6 e 7 ciascuno dei quali specifica una delle tre tipologie di criteri ipotizzate come necessarie ad estrarre una relazione. Il blocco 4 specifica la struttura del *pattern* sintattico che i sintagmi devono avere: il 'criterio sintattico'. Il blocco 6 specifica il contenuto lessicale del nodo centrale del grafo del sintagma: il 'criterio semantico'. Il blocco 7 applica il vincolo di appartenenza dell'oggetto della relazione al contesto il 'criterio di dominio'.

Il 'criterio sintattico' specifica la struttura che i sintagmi devono rispettare sotto forma di grafo di relazioni tra i termini che li compongono. Questa viene formalmente definita utilizzando le proprietà che rappresentano le dipendenze UD. Alla struttura delle dipendenze sintattiche possono essere aggiunti anche criteri morfologici che i termini del sintagma devono rispettare. Nel caso specifico si noti che nel blocco 4 è previsto che il *token* in posizione 2 deve essere annotato come PoS di tipo 'NOUN' (nome).

Il 'criterio semantico', necessario per garantire la coerenza della relazione che verrà asserita col contenuto del sintagma identificato col 'criterio sintattico', viene definita come vincolo di appartenenza ad un insieme controllato di termini di uno o più *token* del sintagma. Formalmente il vincolo è espresso attraverso l'istruzione FILTER che valuta una espressione booleana di tipo *rdfTerm IN* (valore 1, ..., valore n). Nel caso del blocco 6 l'insieme dei valori considerati è limitato al singolo elemento 'figlio'. Questo è legato all'ampiezza della semantica espressa dalla relazione *famTree:isSonOf* che la regola mira ad identificare. La proprietà infatti asserisce una relazione tra un genitore ed un figlio maschio. Sarebbe quindi possibile per aumentare la *recall* estendere l'insieme prevedendo altri termini come "figliolo" o "rampollo", ma non "figlia".

Il criterio di coerenza ai contenuti del dominio applica invece un vincolo espresso al di fuori del contenuto della didascalia del personaggio. A questo livello è possibile prevedere l'elaborazione di qualsiasi criterio purché esprimibile attraverso un costrutto del linguaggio SPARQL. Nel nostro esempio, il blocco 7 utilizza un costrutto del linguaggio che esegue una interrogazione su una base di conoscenza separata per verificare che il personaggio identificato come secondo termine della relazione appartenga alla stessa opera¹²⁰. Si veda nel paragrafo 4.2.5 come questo approccio possa essere efficacemente impiegato per utilizzare contenuti informativi derivati da fonti esterne.

¹²⁰ Nel caso in analisi la funzione attinge al repertorio delle opere del *dataset* Corago LOD. Il grafo di riferimento su cui opera la regola essendo uno strumento intermedio di analisi è tenuto separato dal *dataset* principale. Questo *dataset*, contenente l'annotazione delle descrizioni dei personaggi riferiti all'intero repertorio, è accessibile attraverso l'identificativo DOI: <https://doi.org/10.6084/m9.figshare.7880489>. Vengono fatte confluire nella base di conoscenza solo le triple estratte in base alle regole e validate in base a criteri di valutazione della qualità del dato estratto.

In conclusione, in base all'esempio mostrato, si ipotizza la possibilità di indagare i contenuti descrittivi dei personaggi riportati nelle didascalie attraverso l'uso di procedure di IE in grado di analizzare la struttura (sintassi), i contenuti (lessico) e applicare vincoli di dominio (repertorio e fonti documentali) per produrre informazione qualificata e rappresentabile formalmente.

4.2.4.1 Il caso di studio: le relazioni interpersonali in Metastasio e Goldoni.

La metodologia di estrazione dell'informazione fin qui descritta è stata applicata ad un campione di personaggi derivato dal repertorio delle opere basate sui libretti di Pietro Metastasio¹²¹ e di Carlo Goldoni¹²². La scelta di questi due autori come caso di studio è motivata da una serie di fattori. Il primo è la appartenenza al medesimo periodo storico ovvero la parte centrale del XVIII secolo¹²³, da cui ci si attende una uniformità del lessico utilizzato nelle didascalie. Il secondo è l'inclusione di personaggi derivati sia dal genere comico che dal serio. Il terzo è una numerosità significativa delle opere censite a catalogo: 1122 le opere attribuite a Metastasio, 333 a Goldoni. Nel complesso sono state considerate 1103 opere per un totale di circa 9000 personaggi. Di questi sono stati analizzate circa 6310 didascalie.

La prima analisi ha riguardato l'estrazione delle qualifiche più frequentemente attribuite ai personaggi del campione. Il *pattern* di estrazione ha selezionato i modificatori nominali direttamente riferiti al personaggio. La tabella 4.2 riporta i primi venti lemmi più frequenti nelle qualifiche estratte per i personaggi del repertorio analizzato con la relativa corrispondenza al significato attestato sul *synset* Babelnet (Navigli e Ponzetto, 2012).

Tabella 4.2: Lemmi più frequenti nelle qualifiche del caso di studio

LEMMA QUALIFICA	CONTEGGIO PERSONAGGI	CONCETTO BABELNET
amante	774	http://babelnet.org/rdf/s00052143n
sorella	366	http://babelnet.org/rdf/s00071838n
padre	269	http://babelnet.org/rdf/s00009616n
re	199	http://babelnet.org/rdf/s00049141n
principessa	180	http://babelnet.org/rdf/s00064417n
figlia	103	http://babelnet.org/rdf/s00025364n
amico	86	http://babelnet.org/rdf/s00036538n
principe	82	http://babelnet.org/rdf/s00064405n
moglie	77	http://babelnet.org/rdf/s00053531n
prefetto	69	http://babelnet.org/rdf/s00064128n
nemico	61	http://babelnet.org/rdf/s00001582n
parte	60	
vedova	58	http://babelnet.org/rdf/s00081167n
regina	52	http://babelnet.org/rdf/s00034025n
guardia	51	http://babelnet.org/rdf/s00042043n
tutore	51	http://babelnet.org/rdf/s00025885n
figlio	50	http://babelnet.org/rdf/s00012571n

¹²¹ Riferimento URI dell'autore: <http://corago.unibo.it/resource/RESPONS/Z00000337900>.

¹²² Riferimento URI dell'autore: <http://corago.unibo.it/resource/RESPONS/Z00000244400>.

¹²³ La prima opera censita in archivio di Metastasio è la *Angelica* del 1720 (riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/0000157674>), del Goldoni è *L'uomo di mondo* del 1728 (riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/0001486902>). Le ultime opere del Goldoni risalgono al 1762 un anno prima della sua morte e del Metastasio al 1781.

pretoriano	35	http://babelnet.org/rdf/s00063997n
primogenito	31	http://babelnet.org/rdf/s00030108n

Proprio l'analisi dei termini più frequenti nel campione analizzato ha suggerito di procedere con l'estrazione di relazioni interpersonali tra i personaggi.

In questo caso la procedura di estrazione dovrà ricavare una relazione tra due personaggi applicando una regola basata sul modello descritto nel listato 4.1. Ad esempio: il personaggio Farnaspe¹²⁴ descritto come: "principe parto, amico e tributario d'Osroa, amante e promesso sposo d'Emirena" deve essere messo in relazione con Emirena¹²⁵, descritta come: "prigioniera d'Adriano, amante di Farnaspe". Il risultato atteso dalla procedura di estrazione è una tripla che asserisca: Farnaspe relationship:lifePartnerOf Emirena.

L'analisi effettuata utilizzando la regola di estrazione descritta sul campione dei personaggi di Metastasio e Goldoni ha prodotto l'insieme di relazioni riportate nella tabella 4.3.

Tabella 4.3: Relazioni interpersonali estratte

LEMMA	RELAZIONE	CONTEGGIO
amante	relationship:lifePartnerOf	240
padre	famtree:isFatherOf	79
amico/a	relationship:friendOf	40
madre	famtree:isMotherOf	24
sorella	famtree:isSisterOf	7
fratello	famtree:isBrotherOf	4
figlia	famtree:isDaughterOf	3
zia	famtree:isAuntOf	2
nipote	famtree:isNephewOf	1
moglie	famtree:isWifeOf	1
figlio	famtree:isSonOf	1

Nel complesso la sperimentazione ha evidenziato l'efficacia dell'approccio che prevede la costruzione di regole di estrazione che uniscano criteri sintattici e semantici. Sulla base di questi risultati preliminari è possibile ipotizzare l'evoluzione della procedura attraverso due direttrici. La prima dovrà estendere i *pattern* per essere in grado di cogliere un maggior numero di articolazioni sintattiche utili. La seconda che perfezioni

¹²⁴ Riferimento URI del personaggio: <http://corago.unibo.it/resource/PERSONAG/0000437343>

¹²⁵ Riferimento URI del personaggio: <http://corago.unibo.it/resource/PERSONAG/0000437342>

l'integrazione dei dizionari dei termini per estendere l'analisi ad altre *features* del personaggio utili a popolare in modo analitico i diversi livelli del modello descrittivo proposto per il personaggio d'opera.

4.2.5 Estrazione basata su fonti esterne

La metodologia di estrazione dell'informazione presentata nel paragrafo 4.2.4 prevede l'individuazione di una corrispondenza biunivoca tra la relazione estratta ed il lessico che la qualifica all'interno del testo analizzato. L'individuazione della corrispondenza richiede una preventiva ricerca di un dizionario di termini da associare a ciascuna *feature* che si intende individuare. Nel caso in cui le caratteristiche che si vogliono individuare riguardino entità dotate di nome proprio (*named entity*), è possibile utilizzare dizionari per classificare i termini corrispondenti presenti nel testo. Ad esempio, disponendo di un elenco di nomi di personaggi è possibile annotare tutte le occorrenze di termini corrispondenti ad una 'voce' del dizionario. In ambito NLP, questo metodo è indicato come *Named Entity Recognition* (NER). La procedura di annotazione del testo deve essere dotata di un elenco di 'entità note', ciascuna delle quali con una etichetta associata da utilizzare come *training set* per addestrare lo strumento di annotazione¹²⁶.

L'approccio statistico applicato all'individuazione di entità note è efficace a patto che si sia in grado di costruire un *training set* sufficientemente ampio da coprire adeguatamente le occorrenze presenti nel dominio di interesse. Per l'addestramento si impiegano di norma dizionari preesistenti (Martin e Jurafsky, 2009, p. 768). Nel nostro caso questo metodo non è efficace perché la varietà delle forme e dei nomi da individuare in un testo non è enumerabile a priori.

Nel paragrafo 3.1.1 si è infatti individuata l'opportunità di ricavare dalle ambientazioni delle opere le informazioni utili a collocare l'opera in un contesto geografico. In questo caso la procedura di IE dovrà individuare all'interno del testo che descrive l'ambientazione i termini che rappresentano un toponimo ed associarli alla corrispondente risorsa *linked data* individuata che descrive la località corrispondente. Ad esempio, per l'opera *Chi la fa l'aspetta*¹²⁷ composta da Girolamo Polani su libretto di Francesco Passarini, il libretto reca l'indicazione seguente: "la scena si finge in una contrada di Venezia".

In questo caso la procedura deve essere in grado di individuare il *token* "Venezia" come toponimo e di indicare come risorsa correlata disponibile in rete l'URI: <http://dbpedia.org/page/Venice> corrispondente alla città di Venezia in DBPedia¹²⁸. L'obiettivo della procedura di estrazione dell'informazione dai testi delle ambientazioni è quindi individuare i termini candidati come possibili toponimi e la relativa attestazione su una base di conoscenza alla quale riferire l'identità della località e le relative coordinate geografiche.

Come per l'estrazione delle caratteristiche dei personaggi, anche nel caso dei toponimi si pone il problema della disambiguazione dei riferimenti individuati. Se nel primo caso il problema è strettamente legato alla potenziale polisemia dei termini inclusi nel lemmario di ciascuna proprietà, nel secondo è più semplicemente legato ai casi di omonimia.

Una errata attribuzione di significato ad un termine nel caso dell'estrazione di relazioni tra personaggi porta ad inferire una relazione non attestata nella descrizione. Questo errore può essere mitigato con l'inserimento nel dizionario di riferimento della singola relazione di termini non polisemi, con alone semantico ridotto o la cui disambiguazione sia facilmente desumibile dal contesto sintattico.

Nel caso di omonimie tra località, il toponimo non è facilmente distinguibile e in questo caso il contesto sintattico non aiuta. Per disambiguare le omonimie è necessario considerare vari elementi (o indizi) esterni

¹²⁶ Per una introduzione esaustiva alla problematica dell'estrazione di *named entity* si rinvia a (Martin and Jurafsky, 2009, p. 761).

¹²⁷ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/APC0004467>.

¹²⁸ Per una introduzione alla base di conoscenza DBPedia si può fare riferimento a (Lehmann et al., 2012). Indirizzo del portale di consultazione: <https://wiki.dbpedia.org/> (verificato il 24/03/2019).

alla sola ambientazione. Per esempio, occorre consultare la scheda o il testo completo dell'opera. Anche in considerazione del fatto che i casi rilevati di omonimia sono dell'ordine delle decine di unità si è quindi deciso di procedere con una procedura semi supervisionata.

La procedura di estrazione è stata articolata in tre fasi: annotazione NLP delle ambientazioni delle opere¹²⁹; individuazione dei termini annotati come sostantivi ed interrogazione di DBPedia¹³⁰; creazione delle triple di associazione tra toponimo e URI della risorsa DBPedia¹³¹.

L'annotatore NER è stato addestrato su un set costruito a partire dal testo presente nelle descrizioni di ambientazioni e didascalie dei personaggi¹³². Le due fasi successive hanno semplicemente previsto l'iterazione sull'insieme dei *token* annotati come toponimi per identificare una corrispondenza tra il lemma del *token* e un'entrata DBPedia. L'allineamento tra toponimo e risorsa DBPedia è stato calcolato utilizzando una interrogazione SPARQL il cui template è riportato nel listato 4.2¹³³.

```
#Interrogazione SPARQL per l'estrazione delle corrispondenze tra toponimi e località censite in DBPedia
SELECT DISTINCT ?s
{
  ?s a <http://dbpedia.org/ontology/Place>
  {?s <http://www.w3.org/2000/01/rdf-schema#label> \"\"+toponym+\"\"@it}
  UNION
  {?s <http://www.w3.org/2000/01/rdf-schema#label> \"\"+toponym+\"\"@fr}
  UNION
  {?s <http://www.w3.org/2000/01/rdf-schema#label> \"\"+toponym+\"\"@en}
  UNION
  {?s <http://www.w3.org/2000/01/rdf-schema#label> \"\"+toponym+\"\"@de}
  UNION
  {?s <http://www.w3.org/2000/01/rdf-schema#label> \"\"+toponym+\"\"@es}
}
```

Listato 4.2: Interrogazione SPARQL per l'individuazione corrispondenze toponimi

Nel complesso la procedura di estrazione ha prodotto 4711 associazioni tra *token* annotati provenienti dalle ambientazioni delle opere e risorse DBPedia¹³⁴. La stessa procedura è stata applicata ai toponimi estratti dalle descrizioni di personaggi producendo 7763 associazioni.

I toponimi individuati, le relative associazioni con le entità di provenienza (sia opere che personaggi) e le località DBPedia sono confluite in un unico elenco cui sono stati aggiunti i toponimi ricavati dai luoghi degli spettacoli e di edizione dei libretti.

Su questo elenco complessivo è stata effettuata l'attività di revisione e disambiguazione manuale delle omonimie che ha prodotto un totale di 2808 elementi, di cui 2703 associati a risorse DBPedia.

¹²⁹ Le opere di cui è stata analizzata la ambientazione sono state 6027 (versione dell'archivio al 24/03/2018).

¹³⁰ Per selezionare le sole risorse relative a località geografiche si è verificato che fossero dichiarate appartenere alla classe `dbpedia:Place`.

¹³¹ L'illustrazione dell'algoritmo di estrazione e del relativo codice è riportata nella appendice C.2.

¹³² Si rinvia al paragrafo C.2.1 dell'Appendice C per la descrizione dell'algoritmo di generazione del *training set*.

¹³³ Il servizio per l'interrogazione SPARQL della base di conoscenza DBPedia utilizzato è accessibile all'URL: <http://dbpedia.org/sparql/> (verificato il 23/03/2019). Si noti come il codice della query preveda una corrispondenza letterale tra il toponimo e l'etichetta della risorsa DBPedia di tipo `dbpedia:Place`. Questo ha comportato la necessità di prevedere la gestione della localizzazione delle stringhe pena l'alto numero di mancate corrispondenze individuate.

¹³⁴ Complessivamente gli scarti sono stati 127 per i toponimi derivati dalle ambientazioni e 282 per quelli derivati dalle didascalie dei personaggi. Le cause principali di mancata associazione sono derivate da termini erroneamente annotati come toponimi o a toponimi non presenti in DBPedia come "Pianoro" (ambientazione dell'opera: *Voglia di dote e non di moglie*, riferimento URI: <http://corago.unibo.it/resource/BRANI/0000096052>).

Nelle attività di revisione sono stati trattati in particolare casi come:

- 1) la disambiguazione di omonimie, ad esempio Alessandria (<http://dbpedia.org/resource/Alessandria>) e Alessandria d'Egitto (<http://dbpedia.org/resource/Alexandria>);
- 2) lo schiacciamento di forme diverse su un'unica risorsa DBPedia, ad esempio: Wien, Wien in Österreich, Wienn, Vienna attestati su <http://dbpedia.org/resource/Vienna>.

Nel complesso l'attività di estrazione e di associazione a risorse esterne ha prodotto i risultati riportati nella tabella 4.4.

Tabella 4.4: Toponimi estratti per tipologia di entità

Entità	Toponimi	Associati a risorse DBPedia
Opere	7538	7536
Personaggi	6538	6537
Documenti	51560	51500
Eventi	38541	38435

Sulla base delle relazioni individuate tra opere, eventi e documenti si è potuto procedere all'estrazione delle coordinate geografiche associate alle risorse DBPedia (vedi paragrafo 4.3.1).

4.3 Allineamento ed inclusione di contenuti da fonti LOD

Nel paragrafo 4.2 sono stati illustrati i metodi utilizzati per estendere il patrimonio informativo della base di conoscenza procedendo all'estrazione di informazione strutturata dai contenuti presenti in forma non strutturata all'interno dell'archivio. Una seconda via per espandere la base di conoscenza, forse quella più naturale all'interno dell'ecosistema *linked data* (LD), è l'allineamento¹³⁵ con fonti informative esterne da cui prelevare dati.

L'esplicitazione della semantica del modello concettuale attraverso ontologie formali ha l'obiettivo di rendere interoperabili i contenuti della singola base di conoscenza col patrimonio informativo dell'ecosistema LD. L'interoperabilità basata sulla semantica prevede la possibilità di stabilire legami tra contenuti gestiti da soggetti attraverso sistemi eterogenei per natura e finalità. In questo scenario le ontologie sono i vocabolari fondamentali per accedere al dominio di competenza di ciascuna base di conoscenza (*Knowledge Base* - KB).

Per associare i contenuti di fonti esterne ai propri bisogna individuare un allineamento tra i diversi vocabolari utilizzati. Questo comporta la necessità di individuare corrispondenze sia tra i modelli concettuali sia di singola istanza di entità. Ad esempio, richiamando il caso relativo all'individuazione di toponimi all'interno delle ambientazioni delle opere (vedi paragrafo 4.2.5) si è dovuto individuare sia una corrispondenza tra le classi che descrivono il toponimo nelle due KB, sia la corrispondenza tra le stringhe che identificano una località specifica nei due sistemi.

Nel primo caso si deve identificare una corrispondenza a livello semantico tra i termini che descrivono le classi di entità correlate, nel secondo, dovranno essere definiti i criteri in base ai quali creare la relazione tra le risorse identificate come omologhe.

La natura stessa della relazione che vorrà stabilire tra le due istanze deve essere valutata attentamente. Ad esempio, che tipo di relazione intercorre tra l'appellativo di un toponimo (rappresentato in Corago LOD dalla classe `crm:E44_Place_Appellation`) come "Venezia" e la risorsa DBPedia "Venezia" che ne rappresenta la omonima località¹³⁶?

Le procedure necessarie per importare in Corago LOD le informazioni da fonti *Linked Data* sono state sperimentate attraverso due casi di studio specifici: l'importazione delle coordinate geografiche relative ai toponimi; l'allineamento con le opere contenute nelle fonti presenti nel *Répertoire International des Sources Musicales* (RISM).

4.3.1 Georeferenziazione dei toponimi

Prendiamo in considerazione un esempio di entità di località censita in DBPedia: Ravenna¹³⁷. Come prima considerazione va notato che la classe `dbpedia:place`¹³⁸ è la radice (ovvero la classe archetipo o *ancestor*) della gerarchia di classi specializzate che definiscono in DBPedia le entità di tipo geografico. Nel caso di Ravenna questa risulta essere definita, al massimo grado di specializzazione previsto dall'ontologia, come

¹³⁵ A livello terminologico qui si adotta la terminologia definita in (Euzenat e Shvaiko, 2007, p. 42), con la sola accezione del termine allineamento che qui viene estesa anche alla fase preliminare di analisi delle corrispondenze tra i singoli concetti di due e non solo come risultante di quest'ultima.

¹³⁶ Un interessante studio delle problematiche relative alla rappresentazione dell'identità tra risorse nell'ecosistema Linked Data è presentato in (Halpin et al., 2010).

¹³⁷ Riferimento URI: <http://dbpedia.org/resource/Ravenna>.

¹³⁸ La declaratoria della classe recita: "Immobile things or locations" (riferimento URI alla definizione: <http://dbpedia.org/ontology/Place>).

dbpedia:City (la cui declaratoria riporta: “a relatively large and permanent settlement, particularly a large urban settlement”). Le entità di tipo dbpedia:Place sono definite anche come geo:SpatialThing e vengono annotate con le proprietà geo:lat e geo:long che indicano le relative coordinate geografiche in formato numerico.

La denominazione della località è presente sia come rdfs:label sia come foaf:name. Solo la prima è utilizzata per riportare le diverse forme del nome nelle diverse lingue censite. Non è prevista invece una specifica classe che definisca l'appellativo (ovvero il toponimo) della località come è previsto invece dal CIDOC CRM con la classe crm:E44_Place_Appellation. Come conseguenza, il confronto tra i toponimi presenti in Corago LOD e le località censite in DBPedia avviene sulla base della uguaglianza tra le relative stringhe (vedi listato 4.2).

L'estrazione delle coordinate geografiche della località DBPedia associata ad un toponimo Corago attraverso la proprietà rdfs:seeAlso può essere fatta attraverso l'interrogazione SPARQL del listato 4.3.

```
#Interrogazione SPARQL per l'estrazione delle coordinate geografiche dalle località censite in DBPedia
SELECT
{
  ?locDBPedia <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat.
  ?locDBPedia <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long.
}
WHERE
{
  ?sourceCorago <http://www.w3.org/2000/01/rdf-schema#seeAlso> ?locDBPedia.
  SERVICE <https://dbpedia.org/sparql> {
    ?locDBPedia <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat.
    ?locDBPedia <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long.}
}
```

Listato 4.3: Esempio di estrazione di coordinate geografiche

Una volta associate le coordinate al toponimo di riferimento diventa possibile navigare la relazione tra lo stesso e l'entità da cui è derivato (vedi tabella 4.4 per la numerosità delle relazioni).

Secondo il modello delle proprietà di navigazione (illustrato nel paragrafo 2.5.3), è possibile definire delle proprietà specifiche che associano la località DBPedia alle entità di Corago LOD. Ad esempio, nel caso degli eventi, la proprietà corago:CPC17_performance_coord, assocerà lo spettacolo alla località in cui si è svolto¹³⁹.

Per rispondere ai requisiti di accesso alle entità principali dell'archivio, è necessario che queste vengano indicizzate in base alla collocazione geografica (vedi paragrafo 3.1.1). A questo scopo è stata prevista nell'ontologia Corago SM la proprietà di tipo dato: corago:CDP13_location¹⁴⁰ che permette di specificare le coordinate in base alle quali le risorse dell'archivio vengono individuate nelle ricerche spazio-tempo.

Il risultato della procedura di estrazione descritta ha permesso di arricchire il patrimonio informativo dell'indice con contenuti provenienti da una fonte esterna rendendo possibile l'implementazione di una nuova modalità di accesso e rappresentazione dei suoi contenuti (vedi paragrafo 5.2).

¹³⁹ In questo caso, la proprietà di navigazione rappresenta la forma sintetica del grafo che collega l'istanza di F31_Performance con la relativa DBPedia:Place. La *property path* completa è la seguente: P7_took_place_at/P1_is_identified_by/rdfs:seeAlso.

¹⁴⁰ La proprietà definita come OWL *data property* riporta in nodi di tipo *literal* le coordinate geografiche nel formato previsto dal *repository* per l'impiego in ricerche di tipo *spatial* (<http://www.bigdata.com/rdf/geospatial/literals/v1#lat-lon>).

4.3.2 Associazione Opere al RISM

Il caso di studio riportato nel paragrafo 4.3.1 evidenzia l'utilità dell'allineamento tra basi di conoscenza eterogenee per acquisire informazioni che esulano dal perimetro gestito dal processo di catalogazione dell'archivio. Si pone quindi il tema dell'utilità dell'allineamento rispetto a fonti informative esterne che contengono informazioni contigue ai contenuti dell'archivio. Per sperimentare questa tipologia di allineamento si è ipotizzato di individuare le corrispondenze tra le opere censite in *Corago* le fonti musicali all'interno del RISM.

Il RISM è un catalogo di fonti musicali, manoscritti e musica a stampa con l'obiettivo di "conoscere l'esistente e dove è conservato"¹⁴¹. Complessivamente raccoglie oltre un milione di notizie bibliografiche (di cui circa centosettanta mila sono le edizioni a stampa) di poco più di dieci mila compositori¹⁴². Il sistema OPAC è mantenuto dalla Bayerische Staatsbibliothek di Monaco di Baviera che ha curato anche la pubblicazione della versione in RDF interrogabile anche attraverso un *endpoint* SPARQL¹⁴³.

La sperimentazione ha cercato di individuare in modalità non supervisionata le corrispondenze tra il repertorio Corago e le occorrenze nel RISM. A questo scopo è stato adottato come criterio per identificare istanze omologhe il titolo dell'opera e la corrispondenza delle denominazioni degli autori.

In Corago LOD, per le entità di tipo `frbroo:F1_Work`, i due attributi sono specificati dalla proprietà `rdfs:label` per il titolo dell'opera e dalla relazione con gli autori sintetizzabile attraverso la *property chain*: `corago:CP2_carried_out_role/corago:CP3_carried_out_actor`.

```
#Interrogazione SPARQL per l'estrazione delle opere corrispondenti tra Corago LOD e RISM
SELECT DISTINCT ?workCorago ?workRism {
  #BLOCCO 0: estrazione delle opere (frbroo:F1_Work) e dei relative autori in Corago LOD
  ? workCorago a <h workCorago ttp://erlangen-crm.org/efrbroo/F1_Work>.
  ? workCorago rdfs:label ?wLbl.
  ? workCorago <http://erlangen-crm.org/efrbroo/R16i_was_initiated_by> ?wCon.
  ?wCon <http://corago.unibo.it/sm/CP2_carried_out_role>/<http://corago.unibo.it/sm/CP3_carried_out_actor> ?wAuth.
  ?wAuth rdfs:label ?wAuthLbl.
  #BLOCCO 1: estrazione mediante query federata delle opere e dei relativi autori in RISM
  SERVICE <https://data.rism.info/sparql>
  {
    ? workRism <http://purl.org/dc/terms/title> ?wRismTit.
    ? workRism <http://purl.org/dc/terms/creator> ?wRismAuth.
    ?wRismAuth <http://xmlns.com/foaf/0.1/name> ?wAuthLbl.
  }
  #BLOCCO 2: applicazione dei criteri di match tra le istanze individuate nei due sistemi
  FILTER (CONTAINS(STR(?wRismTit),STR(?wLbl)) AND (REGEX(STR(?wAuthLbl),STR(?wAuthLbl),'i'))))
}
```

Listato 4.4: Estrazione delle opere corrispondenti tra Corago LOD e RISM

La trasposizione in formato RDF della scheda dell'opera nel RISM è ottenuta con l'adozione del vocabolario DCMI Metadata Terms¹⁴⁴. In questo formato il titolo dell'opera è espresso attraverso la proprietà:

¹⁴¹ Citato nella presentazione istituzionale online: <http://www.rism.info/organisation.html> (verificato il 31/03/2019).

¹⁴² Dati ricavati dall'Annual Report 2018, accessibile online all'indirizzo: <http://www.rism.info/publications/annual-reports/2018.html> (verificato il 31/03/2019).

¹⁴³ Il servizio è accessibile all'indirizzo: <https://data.rism.info/sparql> (verificato il 31/03/2019).

¹⁴⁴ Una introduzione al formato e alla sua implementazione è reperibile all'indirizzo: <http://dublincore.org/schemas/rdfs/> (verificato il 31/03/2019).

dcterms:title; gli autori sono associati all'opera mediante la proprietà: dcterms:creator ed i rispettivi nomi sono specificati con la proprietà: foaf:name¹⁴⁵.

Il listato 4.4 riporta un esempio di interrogazione eseguibile come query federata sui servizi SPARQL dei due sistemi. Nel primo blocco (blocco 0) vengono selezionate tutte le opere (entità di tipo frbroo:F1_Work) in Coago LOD e l'elenco dei relativi autori. Nel blocco 1 vengono estratte le stesse informazioni dal catalogo RISM. Nell'ultimo blocco (blocco 2) vengono applicati i criteri utili ad identificare le istanze provenienti dai due sistemi che condividono titolo e autori. A questo livello possono essere implementate le eventuali euristiche per migliorare l'indice di *recall* della procedura. Nell'esempio proposto sono adottati due criteri di corrispondenza diversi tra titoli e autori. Per i titoli il criterio specifica che il titolo dell'opera ricavata dal RISM deve contenere il titolo dell'opera in Corago LOD (criterio implementato mediante la funzione SPARQL: CONTAINS). Per i nomi degli autori invece è previsto un semplice confronto tra stringhe di tipo *case insensitive* (implementato attraverso la funzione REGEX per la gestione di espressioni regolari).

Utilizzando la procedura descritta sono state individuate 340 corrispondenze. I riferimenti così ottenuti sono stati integrati nel *dataset* Corago come riferimenti espressi mediante la proprietà rdfs:seeAlso¹⁴⁶.

Nel complesso la sperimentazione evidenzia come l'interoperabilità tra basi di conoscenza afferenti allo stesso ambito possa avvenire sulla base di euristiche non supervisionate a patto di poter individuare un allineamento stretto tra i rispettivi modelli concettuali di riferimento e una significativa sovrapposizione di contenuti¹⁴⁷. La possibilità di stabilire corrispondenze corrette in modo non supervisionato è un passaggio necessario per rendere realmente percorribile una catalogazione distribuita all'interno dell'ecosistema Linked Data (Bonora e Pompilio, 2019).

¹⁴⁵ L'authority file degli autori del RISM prevede l'utilizzo della forma canonica [cognome], [nome].

¹⁴⁶ In questo caso, trattandosi di risultati ottenuti mediante una procedura non supervisionata si è preferito esprimere la relazione di corrispondenza con la proprietà generica di rinvio tra entità rdfs:seeAlso piuttosto che con la più stringente proprietà rdfs:sameAs che indica la perfetta identità tra le due risorse collegate. Per un approfondimento sul tema dell'identità tra risorse LOD si rinvia a (Halpin et al., 2010).

¹⁴⁷ Per sua natura il RISM del repertorio operistico contiene prevalentemente arie ed estratti pubblicati in miscellanee a stampa, ridotto è invece il numero di libretti. Questo giustifica il risultato numericamente limitato di corrispondenze trovate.

5 Strategie di accesso ai contenuti

L'obiettivo finale della sperimentazione è rendere accessibile alla comunità di riferimento il patrimonio informativo dell'archivio *Corago* nella sua nuova forma. A questo scopo è stata realizzata una applicazione di consultazione del *dataset* Corago LOD. La diponibilità della nuova applicazione ha consentito al tempo stesso di valutare pregi e difetti della nuova soluzione rispetto a quella tradizionale¹⁴⁸. Il sistema è stato disegnato a partire dai requisiti funzionali descritti nel paragrafo 1.5. In particolare, sono state implementate le funzionalità di ricerca dei contenuti (requisiti [REQ1], [REQ4], [REQ5] e [REQ6]) e quelle necessarie per la rappresentazione delle diverse tipologie di entità (requisiti [REQ2] e [REQ3]).

Le funzionalità di accesso ai contenuti dell'archivio sono: la ricerca libera testuale, la ricerca spazio-tempo, la ricerca avanzata e ricerca per personaggi. Sono state inoltre predisposte: le pagine di presentazione dei risultati delle diverse tipologie di ricerca, la pagina di visualizzazione delle schede delle risorse e la pagina di visualizzazione dati su cartografia. Nei paragrafi seguenti vengono illustrate le caratteristiche delle diverse funzioni di ricerca e delle interfacce di presentazione dei relativi risultati.

Per realizzare gli strumenti di consultazione sono stati considerati profili utente diversi: dall'utente esperto del dominio ed informato rispetto al modello FRBR, al semplice utente generico interessato alla materia. Al fine di valutare il sistema realizzato è stato effettuato un rilevamento del livello di gradimento attraverso la somministrazione di un questionario.

5.1 Ricerche testuali

Nel realizzare funzionalità di ricerca che impiegano criteri testuali per individuare entità descritte attraverso modelli strutturati (ricerche *entity-oriented*) si devono considerare due fattori che le distinguono in modo significativo dalle ricerche orientate ai documenti (Bast et al., 2012). Il primo fattore è la maggiore rilevanza della *recall* rispetto alla *precision*: è preferibile infatti includere risultati potenzialmente 'rumorosi' piuttosto che escluderne sistematicamente alcuni utili. L'esclusione sistematica di risultati deriva dalla capacità del sistema di considerare un insieme sufficientemente ampio di attributi testuali del modello di ciascuna entità per riuscire ad intercettarne tutte quelle caratteristiche che rientrano tra i criteri utilizzati nelle diverse possibili strategie di ricerca adottate dagli utenti. Da qui deriva la necessità di costruire un indice in cui confluiscono quei contenuti testuali che meglio descrivono la singola entità nella prospettiva dell'utente finale. Il secondo fattore è legato all'usabilità della procedura di ricerca ed alla trasparenza dei criteri impiegati per produrre i risultati. In quest'ottica alla semplicità di inserimento dei criteri deve corrispondere una linearità nella loro applicazione. La procedura di consultazione è stata progettata per consentire all'utente di valutare in autonomia l'aderenza dei risultati rispetto ai criteri utilizzati, e quindi in ultima istanza, di stimare la capacità del sistema di individuare i contenuti più rilevanti presenti nella base di conoscenza.

Per implementare i requisiti [REQ1] e [REQ4] i contenuti testuali che compongono l'indice di una risorsa vengono specificati nel modello descrittivo per ciascuna delle tre entità di base dell'archivio: opere, eventi e documenti (vedi paragrafo 2.6.2.2). Ad esempio, potendo scegliere quali attributi sono inclusi nell'indice full-text di un'opera, si può decidere se tra i toponimi utili ad individuarla siano incluse: tutte le località in cui l'opera è stata rappresentata o la sola località della prima assoluta. Contestualmente si potranno escludere le località di nascita o morte degli autori.

¹⁴⁸ Per una descrizione generale dell'architettura del sistema sperimentale si rinvia alla Appendice B.

La possibilità di configurare la composizione dell'indice e di mantenere traccia della relazione col contenuto indicizzato (in base a cui viene individuata l'occorrenza e calcolata la rilevanza del risultato) risponde all'esigenza di rendere trasparente la funzionalità di ricerca.

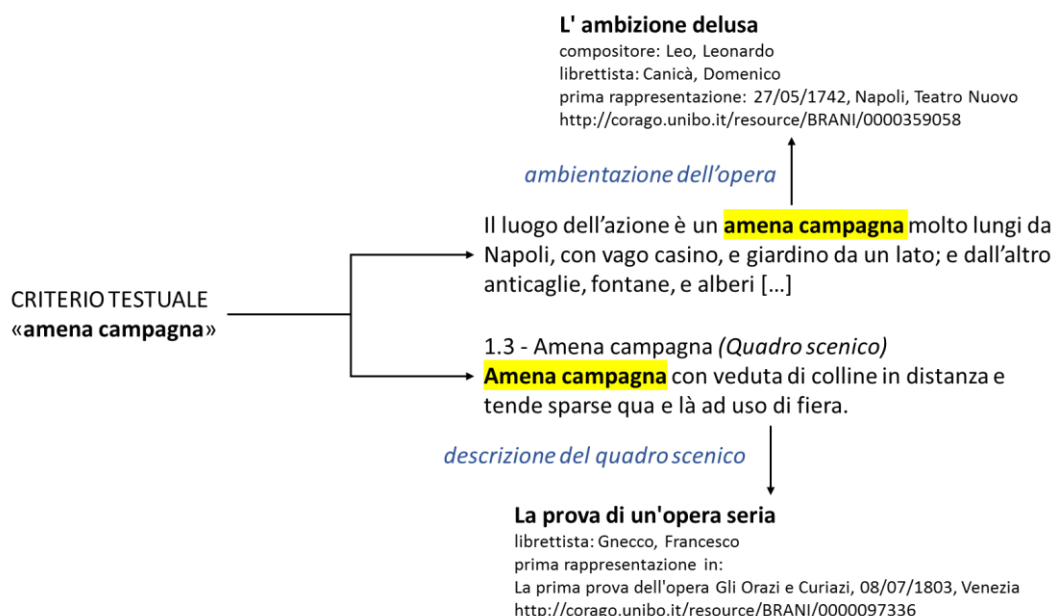


Figura 5-1: Relazione tra criteri di ricerca, contenuti testuali e risultati

La figura 5.1 illustra le corrispondenze tra criterio testuale utilizzato per la ricerca full-text, testi inseriti nell'indice e risultati individuati (in questo caso due opere). La logica di indicizzazione e la conseguente morfologia dell'indice (vedi paragrafo 2.6.2.3) permettono di restituire all'utente assieme al risultato sia il corrispondente testo individuato, sia la semantica della loro relazione. L'esplicitazione del nesso tra testo e risultato consente una verifica immediata della congruità col criterio utilizzato e di valutarne efficacia e rilevanza rispetto agli obiettivi della ricerca.

Risultato	* Responsabili	Data	Tipo	Attributi	Riferimenti
L' *ambizione delusa 📄	compositore: Leo, Leonardo librettista: Canicà, Domenico	27/05/1742	Opera	macrogenere: melodramma genere in fonte: commedia pastorale prima rappresentazione: L' *ambizione delusa lingua: italiano	" Il luogo dell'azione è un amena campagna molto lungi da Napoli, ..." (Ambientazione)

Figura 5-2: Presentazione del risultato di una ricerca full-text

Il sistema colloca quindi il risultato all'interno del 'contesto informativo', riportato nella colonna "Riferimenti" (vedi figura 5.2). Il contesto presenta i criteri in base ai quali il risultato è stato selezionato e va a completare l'insieme dei relativi attributi fondamentali (vedi paragrafo 2.4). In questo modo viene offerta una rappresentazione sintetica sia delle caratteristiche corrispondenti ai criteri di ricerca che di quelle identitarie di ciascun risultato. Il sistema risponde così al requisito [REQ4] che richiede la realizzazione di una funzionalità di ricerca dell'archivio in grado di individuare le entità in base ai rispettivi contenuti testuali offrendo all'utente finale gli strumenti essenziali per valutare la pertinenza dei risultati ottenuti¹⁴⁹.

¹⁴⁹ Si rinvia alla descrizione del questionario di valutazione dell'efficacia del sistema presentato nel paragrafo 5.5 ed in particolare alla discussione dei risultati relativi al quesito Q 2.2 (vedi paragrafo 5.5.1.1) volta a stabilire l'efficacia della rappresentazione dei riferimenti al 'contesto informativo' dei risultati delle ricerche testuali.

Dal punto di vista funzionale il sistema presenta una maschera per la ricerca libera testuale che consente di specificare i seguenti criteri¹⁵⁰: una stringa di testo, l'arco temporale della ricerca e i tipi di entità che saranno oggetto della ricerca¹⁵¹ (figura 5.3).

Ricerca testuale libera:

Arco temporale della ricerca: **1600 - 1950**

1500 201

Tipo di entità: ☒ Opere ☒ Eventi ☒ Documenti

Figura 5-3: Maschera per la ricerca libera testuale

L'utente ha la possibilità di avviare la ricerca indicando come unico criterio quello testuale inserendo uno o più termini nel campo "Ricerca libera testuale". Il sistema precompila i criteri secondari relativi al *range* temporale e alle tipologie di entità nel modo meno restrittivo per far sì che il criterio testuale risulti quello prevalente. L'obiettivo in questo caso è riproporre l'approccio adottato dai motori di ricerca generalisti che riducono all'essenziale la conoscenza relativa al funzionamento del sistema da parte dell'utente¹⁵² necessaria per ottenere risultati utili.

Seguendo questa logica, il sistema considera implicitamente l'inserimento di più termini come criteri congiunti (operatore logico AND) da applicarsi all'intero contenuto testuale indicizzato per ciascuna risorsa. Ad esempio, la stringa "amena campagna" viene ricercata come co-occorrenza dei termini "amena" AND "campagna". Resta comunque possibile specificare l'impiego dei due termini come unica stringa, racchiudendo i termini tra doppi apici: in questo caso il sistema ricerca la forma "amena campagna" (stringa 'esatta'), ovvero considera solo le occorrenze in cui i due termini appaiono contigui e coincidenti.

Dal punto di vista funzionale è opportuno sottolineare due conseguenze derivanti dalla peculiare morfologia dell'indice testuale implementato nel sistema. La prima deriva dal fatto che la ricerca multi-termine considera tutto il testo indicizzato per una risorsa, quindi i diversi termini posso comparire in attributi diversi della scheda dell'entità, ovvero distanza tra i termini e la natura della relazione sono ininfluenti. Ad esempio, utilizzando come criterio i due termini "Zenobia" e "Palmira", il sistema individuerà tra i risultati utili, insieme alle molte "Zenobia in Palmira", anche *Il vinto trionfante del vincitore*¹⁵³ includendo l'opera tra i risultati in base alla presenza tra i personaggi di una "Zenobia" ed un "Ormonte" descritto come "governator di Palmira". Questo esempio evidenzia come il sistema sia in grado di ignorare la relazione tra contenuto e risorsa in fase di ricerca per riportarla invece in modo esplicito in fase di presentazione dei risultati.

¹⁵⁰ La funzione è accessibile all'indirizzo: <http://coragolod1.ing.unibo.it:8080/corago/remoteSearch> (verificata il 10/05/2019).

¹⁵¹ È opportuno ricordare che la struttura dell'indice implementato per la ricerca testuale prescinde dalla natura della risorsa indicizzata: potenzialmente qualsiasi tipo di entità del modello concettuale può essere indicizzato in base ai propri 'FT *breadcrumbs*' (vedi paragrafo 2.6.2). Si tratta quindi di una scelta fatta in fase di impostazione dell'indice a livello concettuale rispetto alla quale, offrire all'utente la possibilità di filtrare i risultati per tipologia, è una scelta di tipo funzionale dell'applicazione di consultazione.

¹⁵² Si fa riferimento alle logiche di funzionamento di base dei principali motori di ricerca generalisti (Spink and Zimmer, 2008), così come dei cosiddetti motori semantici (Sudeepthi et al., 2012).

¹⁵³ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/APC0003933>.

La seconda caratteristica deriva dalla presenza nell'indice degli URI di tutte le entità associate alla risorsa indicizzata. Questa presenza consente di specificare come criterio di ricerca testuale anche il riferimento ad una specifica opera. Ad esempio, se volessi ricercare una edizione in lingua spagnola della *Giulietta e Romeo* di Felice Romani potrei inserire il relativo URI¹⁵⁴ e l'indicazione della lingua¹⁵⁵.

La funzione di ricerca testuale è disegnata per produrre l'insieme più ampio possibile di risultati corrispondenti alle tipologie di risultato selezionate, lasciando all'utente la valutazione della rilevanza dei singoli risultati in funzione del relativo 'contesto informativo'. All'utente viene quindi data la possibilità di raffinare la ricerca operando sui criteri testuali o restringendo i parametri aggiuntivi: arco temporale e tipologie di entità.

Questa apparentemente limitata capacità di controllo sul campo di applicazione dei criteri testuali, ovvero l'impossibilità di specificare su quale attributo del modello debba essere applicato il criterio¹⁵⁶, ha come obiettivo quello di favorire la cosiddetta *serendipity*¹⁵⁷. La 'ricerca libera testuale' è infatti disegnata per offrire un accesso di primo livello rivolto ad una utenza generalista o ad un profilo più esperto che voglia effettuare una ricognizione iniziale dei contenuti presenti nell'archivio. In questo senso il confronto con le funzionalità di ricerca offerte dal sistema di consultazione del sistema tradizionale può essere indicativo¹⁵⁸. Nel sistema tradizionale l'accesso ai contenuti è suddiviso per tipologia di entità: opere, eventi e documenti. L'utente deve quindi preventivamente scegliere per quale tipologia di entità effettuare la ricerca: per ciascuna di esse è infatti prevista una maschera che presenta campi specifici. Per quanto le tre maschere possano risultare semplici ed intuitive, all'utente è comunque richiesto di confrontarsi con l'articolazione del modello descrittivo previsto per le tre entità. La scelta della tipologia da cui avviare la consultazione e come applicare i criteri di filtro risulta inoltre vincolante rispetto al tipo di risultati ottenibili. Ad esempio, utilizzando come criterio per il titolo delle opere "Adriano in Siria" il sistema tradizionale individua 59 opere¹⁵⁹, il sistema Corago LOD ne rileva 91¹⁶⁰. Tra i risultati presentati da entrambi i sistemi compare l'*Adriano in Siria*¹⁶¹ su libretto del Metastasio ma non *Mas gloria es triunfar de sí, Adriano en Syria*¹⁶² che viene invece individuato solo dal nuovo sistema in quanto opera tratta dalla prima¹⁶³.

In modo coerente all'approccio utilizzato per l'indicizzazione dei contenuti testuali, anche l'applicazione del vincolo temporale avviene generalizzando le diverse relazioni che intercorrono tra periodo oggetto della ricerca e risultati. Contrariamente a quanto previsto dalle funzionalità di ricerca del sistema tradizionale (nel quale è necessario indicare espressamente la data di prima esecuzione delle opere, la data degli eventi e

¹⁵⁴ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/Z000053396>.

¹⁵⁵ Nello specifico l'unica occorrenza è la *Julietta y Romeo*, riferimento URI del libretto: <http://corago.unibo.it/resource/DOCUMENTI/0000506791>.

¹⁵⁶ Si noti come questa tipologia di ricerca sia implementata nel nuovo sistema attraverso la funzione di 'ricerca avanzata', descritta nel paragrafo 5.3.

¹⁵⁷ Tra le diverse accezioni del termine in ambito informatico si considera qui quella che la descrive come l'occasionale scoperta di qualcosa di utile. Per una analisi del ruolo giocato dalla *serendipity* nell'accesso alla conoscenza e nell'implementazione dei sistemi informativi che la gestiscono si rinvia a (Race, 2012).

¹⁵⁸ Il sistema è accessibile all'indirizzo: <http://corago.unibo.it> (consultato il 11/05/2019). Riferimento DOI: 10.6092/UNIBO/CORAGO.

¹⁵⁹ Ricerca effettuata sull'aggiornamento dell'archivio al 15 gennaio 2019.

¹⁶⁰ Ricerca effettuata sull'aggiornamento dati dell'archivio al 4 marzo 2018.

¹⁶¹ Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/0000052488>. Riferimento alla scheda nel Corago tradizionale: <http://corago.unibo.it/opera/0000052488>.

¹⁶² Riferimento URI dell'opera: <http://corago.unibo.it/resource/BRANI/0001297340>.

¹⁶³ Nel caso specifico l'*Adriano in Siria* è stato incluso come titolo di opera collegata nell'indice della *Mas gloria es triunfar de sí, Adriano en Syria* in base alla proprietà corago:CPC82_work_ancestors.

l'anno di pubblicazione dei documenti), il criterio di filtro temporale del nuovo sistema prevede la possibilità di operare sull'unica coordinata temporale alla quale vengono ricondotte tutte le relazioni temporali previste dal modello concettuale. Ad esempio, per le opere può essere considerato, oltre alla data di prima esecuzione, anche l'anno di composizione. È inoltre possibile estendere il perimetro di datazione di un'opera fino all'arco di vita dei rispettivi autori o edizione dei libretti che la contengono. Questo vale per tutte le risorse indicizzate e rientra tra le scelte implementative delle modalità di applicazione del criterio stesso.

Nel paragrafo 2.6.2.5 è stata illustrata la modalità di utilizzo generico dell'indice testuale proposto. Per implementare la funzione di ricerca nei termini appena descritti è necessario concatenare i tre parametri previsti: i criteri testuali, l'arco temporale d'interesse e le restrizioni sulle tipologie di risultato.

Template di query SPARQL per la ricerca libera testuale

#BLOCCO 0: selezione dei risultati

```
SELECT distinct ?s (<http://corago.unibo.it/sm/CPA4_ft_breadcrumb> as ?p) ?o {
```

#BLOCCO 1: utilizzo dell'indice full-text

```
?s corago:CPA5_ft_indexed_by ?text.
?text bds:search "Otello Verdi".
?text bds:relevance ?relevance.
```

#BLOCCO 2: applicazione vincolo temporale

```
?t a <http://erlangen-crm.org/current/E49_Time_Appellation>.
?t rdf:value ?tl.
FILTER(?tl >= " + fromTime + "^^xsd:dateTime && ?tl < " + toTime + "^^xsd:dateTime).
?d ecrm:P78_is_identified_by/ecrm:P78_is_identified_by* ?t.
?s

    # insieme di relazioni temporali ammesse
    (corago:CP9_first_performed_in/ecrm:P4_has_time-span) |
    (ecrm:P4_has_time-span) |
    (frbroo:R16i_was_initiated_by/ecrm:P4_has_time-span) |
    (frbroo:CLR6_should_carry/^ecrm:P94_has_created/ecrm:P4_has_time-span) |
    (^ecrm:P67_refers_to)

?d.
```

#BLOCCO 3: restrizione su tipi di risultato

```
?s a ?class.
FILTER (?class IN ( [INSIEME DI TIPI RICHIESTI])
```

#BLOCCHI OPZIONALI

#BLOCCO 4: configurazione parametri utilizzo dell'indice full-text

```
?text bds:matchAllTerms "true" #il parametro specifica la concatenazione in AND dei termini
```

#BLOCCO 5: applicazione criterio di restrizione a stringa esatta

```
FILTER regex(?text,"Otello Verdi","i").
```

Listato 5.1: Template di query SPARQL per la ricerca libera testuale

Il listato 5.1 riporta un esempio semplificato di interrogazione in linguaggio SPARQL impiegata dalla funzione di ricerca. Il primo criterio viene applicato direttamente all'indice selezionando i risultati in base al costrutto riportato nel blocco 1. Il secondo opera specificando le relazioni che devono intercorrere tra i risultati e la dimensione temporale (blocco 2). Il terzo filtra i risultati in base alla relativa tipologia (blocco 3). Seguono alcuni esempi di blocchi opzionali per la configurazione dei requisiti del tipo di concatenazione logica dei

termini e dell'eventuale applicazione della corrispondenza esatta. Il blocco 6 riporta infine un esempio di criterio di ordinamento basato sull'indice di rilevanza assoluta calcolata dalla libreria Lucene (vedi paragrafo 2.6.2).

Occorre osservare come sia possibile applicare criteri di accesso che operino a diversi livelli in unico costruito. È infatti possibile operare direttamente sulle definizioni dei contenuti sfruttando la capacità espressiva del modello concettuale, anziché basarsi sulla articolazione fisica del dato come deve necessariamente avvenire nell'interrogazione di sistemi basati su modelli relazionali. Ad esempio, la restrizione sulle tipologie di risultati è applicata indistintamente a tutti i risultati, nell'approccio relazionale richiederebbe la normalizzazione della tassonomia delle entità. Nel caso del modello dell'archivio tradizionale, in assenza di quest'ultima, il vincolo dovrebbe operare su tre tabelle distinte¹⁶⁴. Allo stesso modo, l'insieme di attributi presi in considerazione per l'applicazione del vincolo temporale è specificato utilizzando le rispettive definizioni. Questo rende possibile specificare in modo analitico quali date verranno considerate e quali escluse, oppure generalizzare includendo indistintamente una qualsiasi relazione di tipo temporale del risultato col periodo di interesse. Lo stesso livello di astrazione è stato raggiunto con l'introduzione del particolare indice full-text che permette di effettuare ricerche *entity-based* partendo da un insieme di contenuti testuali organizzati sotto forma di 'contesto informativo' allineato al modello concettuale sottostante. L'insieme di queste caratteristiche ha portato ad una modalità di accesso che unisce l'immediatezza degli strumenti di ricerca testuali all'analiticità dell'informazione prevista dal modello concettuale di riferimento.

5.2 Ricerche per dimensioni spazio-tempo e di dominio

5.2.1 Ricerche spazio-tempo

Nel paragrafo 3.1.1 viene presentato l'insieme delle informazioni relative alla collocazione geografica delle entità contenute nell'archivio. La localizzazione geografica della produzione teatrale consente di svolgere indagini storiche rilevanti per la storia materiale del melodramma. Ad esempio, è possibile ripercorrere cronologicamente la carriera di un cantante o seguire la fortuna di un'opera rappresentata nei diversi teatri.

La conoscenza del luogo geografico di ambientazione del dramma collegata al luogo ed edificio teatrale nel quale l'opera è stata rappresentata consente di indagare altre prospettive storiche del melodramma come la scelta dei soggetti e la fortuna di alcuni temi in relazione alla geografia culturale nelle diverse epoche della storia del melodramma.

Per rispondere a questo tipo di indagini ([REQ5], vedi paragrafo 1.5.3), il sistema di ricerca deve proporre spazio e tempo come criteri di accesso comuni a opere, eventi e documenti. In questo modo è possibile indagare la co-occorrenza di relazioni tra repertorio, cronologia e fonti documentali. Se la dimensione tempo può essere gestita come un arco temporale, ovvero dando la possibilità di indicare una data di inizio e una data di fine del periodo, la definizione dell'area geografica entro la quale effettuare la ricerca richiede l'impiego della cartografia come strumento per delimitare il perimetro di interesse. La cartografia come metafora visuale offre infatti all'utente uno strumento intuitivo per indicare in modo rapido e preciso la geometria dell'area interessata dalla ricerca e di visualizzarne i risultati in modo intuitivo¹⁶⁵.

¹⁶⁴ Si rinvia all'appendice A.11 per il diagramma relativo alla struttura dati del modello relazionale Corago.

¹⁶⁵ La valutazione circa l'efficacia della cartografia come metafora per l'interazione col sistema è oggetto dei quesiti 1.5 e 2.5 del questionario di valutazione del sistema (vedi paragrafo 5.5).

Filtro sul criterio di localizzazione delle entità: ☒ Ambientazioni delle opere ☒ Località degli eventi ☒ Luoghi di edizione dei documenti

Arco temporale della ricerca: **1500 - 1950**

2000 a.c. 2000 d.c.

1499 1949

Tipo di riferimento temporale: qualsiasi riferimento

Ulteriore criterio di filtro su contenuti testuali:

Figura 5-4: Maschera per la ricerca spazio-tempo

La figura 5.4 rappresenta la maschera realizzata per l’inserimento dei parametri per la ricerca spazio-tempo. La maschera è articolata in tre sezioni principali: la prima è dedicata alla definizione del criterio relativo allo spazio, la successiva all’inserimento dell’arco temporale e in ultimo viene proposto un campo testuale per l’eventuale indicazione di un ulteriore criterio che operi sull’indice full-text.

La definizione dell’area geografica su cui deve essere effettuata la ricerca avviene attraverso il posizionamento sulla mappa dell’area circolare evidenziata in rosso¹⁶⁶. L’utente può posizionare la circonferenza sulla zona desiderata e variarne il diametro utilizzando il mouse. Allo stesso modo è possibile spostarsi liberamente sulla mappa (effettuando il *pan*) ed utilizzando i pulsanti “+” e “-” nell’angolo superiore sinistro adeguare il livello di dettaglio (effettuando lo zoom della mappa). Attraverso la combinazione del diametro dell’area interessata e del livello di zoom della mappa l’utente può circoscrivere la ricerca ad una singola località (diametro di circa 5 km) o estenderla ad un intero continente¹⁶⁷.

¹⁶⁶ Il sistema utilizza la cartografia messa a disposizione dal progetto open source: Open Street Map (Haklay and Weber, 2008). Sito web del progetto: <http://openstreetmap.org> (visitato il 17/05/2019).

¹⁶⁷ In realtà il sistema non impone vincoli specifici: gli esempi riportati si riferiscono a criteri funzionali oltre i quali il criterio impostato diventa superiore alla risoluzione del dato (si veda il paragrafo 4.3.1) o scarsamente selettivo (eccessiva numerosità dei risultati prodotti).

Come ulteriore parametro per la definizione del criterio geografico, il sistema prevede la possibilità di specificare il tipo di relazione che deve intercorrere tra le entità e le rispettive località. Il parametro permette di indicare tre tipi di relazione: l'ambientazione delle opere, la località in cui si sono svolti gli eventi e i luoghi di edizione dei documenti. Implicitamente questo parametro permette all'utente di specificare il tipo di risultato atteso: le opere nel primo caso, eventi e documenti rispettivamente negli altri due¹⁶⁸.

L'indicazione relativa all'arco temporale è gestita attraverso un controllo utente sia di tipo visuale (una *slidebar* permette di definirne i due estremi utilizzando dei cursori) che testuale (due caselle di testo consentono di specificarne le date esatte). Anche in questo caso è previsto un ulteriore parametro attraverso cui specificare quale tipo di relazione con le opere dovrà essere considerata per l'applicazione del filtro. I valori previsti sono: un qualsiasi riferimento temporale, la data di composizione o della prima rappresentazione o l'ambientazione temporale dell'opera. La finalità di quest'ultimo è consentire la ricerca nel repertorio in base al criterio storico o al contenuto. Come ulteriore criterio selettivo, il sistema prevede l'inserimento di un testo che insiste sull'indice full-text del sistema (vedi paragrafo 2.6.1 e 5.1).

Attraverso la composizione dei tre criteri il sistema è in grado di rispondere a quesiti come: "quali opere composte nella prima metà dell'Ottocento sono ambientate in Medio Oriente?", oppure "in quali piazze del nord Italia è stato rappresentato *l'Achille in Sciro*?" o piuttosto individuare le opere rappresentate nel Lombardo Veneto nel 1848. Il sistema presenta i risultati delle ricerche effettuate in base a questi criteri nella forma tabellare comune a tutte le ricerche del sistema (vedi paragrafo 5.4.1) che, a sua volta, prevede la possibilità di riportare le occorrenze individuate sulla cartografia attraverso la relativa funzione (vedi paragrafo 5.4.3).

5.2.2 Ricerche basate sui personaggi

Nel paragrafo 3.2 è stata illustrata la sperimentazione relativa all'impiego del personaggio come dimensione di analisi delle relazioni tra i testi letterari del teatro d'opera. In particolare, nel paragrafo 3.2.3 sono stati presentati i risultati ottenuti attraverso le indagini rese possibili dall'introduzione del concetto di 'figura drammatica' intesa come il primo livello di aggregazione dell'insieme di caratteristiche descrittive comuni a più personaggi. Operativamente le indagini sui contenuti dell'archivio sono state condotte attraverso la funzione di 'ricerca dei personaggi'¹⁶⁹.

¹⁶⁸ Questo criterio di filtro dei risultati corrisponde in termini funzionali al filtro che opera direttamente sulla tipologia di entità nella ricerca libera testuale (vedi paragrafo 5.1).

¹⁶⁹ La funzione è accessibile all'URL: <http://coragolod1.ing.unibo.it:8080/corago/analyticalSearch> (verificato il 18/05/2019).

Ricerca delle entità in base a criteri analitici

Ricerca opere per Figure Drammatiche o loro descrizioni

Nota: è possibile indicare più nomi di figure drammatiche che verranno considerate in AND: il sistema cercherà le opere in cui risultino presenti tutte le figure indicate. E' possibile ricercare sia le denominazioni delle figure drammatiche che le descrizioni complete o parziali.

Avvia la ricerca

Ricerca Figure Drammatiche e relazioni coi personaggi ed opere

Nome	<input type="text"/>		
Descrizione	<input type="text"/>		
Arco temporale opere	da: <input type="text" value="1/1/1500"/>	a: <input type="text" value="31/12/2000"/>	

Nota: è possibile indicare più nomi di figure drammatiche che verranno considerate in AND: il sistema cercherà le opere in cui risultino presenti tutte le figure indicate. E' possibile ricercare sia le denominazioni delle figure drammatiche che le descrizioni complete o parziali.

Avvia la ricerca

Figura 5-5: Maschera di ricerca delle figure drammatiche

La maschera si compone di due sezioni distinte: la prima dedicata alla ricerca puntuale delle opere che contengano il nome della figura o delle figure drammatiche inserite come criteri; la seconda invece permette di ricercare le figure drammatiche in base al nome, ai termini che compaiano nelle didascalie e l'arco temporale di riferimento delle opere cui sono associate. L'obiettivo delle due funzioni è individuare l'insieme di opere che condividono un sottoinsieme di personaggi associati alle figure drammatiche indicate come criterio di ricerca¹⁷⁰.

La maschera utilizza una funzione basata sull'indice delle dimensioni analitiche descritto nel paragrafo 3.2.2.1. L'indice permette di effettuare ricerche full-text su nomi e descrizioni di personaggi e di figure drammatiche associate alle opere. Il risultato della funzione è un elenco di opere che condividono l'insieme di personaggi specificato attraverso i parametri di ricerca. Ad esempio, utilizzando la prima sezione della maschera, se si ricercano i personaggi: Dorimene, Lucilla e Nespola, il sistema individua due opere: *Lo spirito di contraddizione* e *L'ubbidienza per astuzia* che costituiscono il caso di studio presentato nel paragrafo 3.2.3. Se invece si utilizza la seconda sezione per ricercare tutte le "fate" presenti in archivio, si scopre che sono circa ventinove, tra cui compaiono Falsirena e Logistilla protagoniste del Caso 1 riportato nel paragrafo 3.2.3.1. Mentre la prima modalità di ricerca presenta i risultati attraverso la rappresentazione tabellare descritti in 5.4.1, la seconda rinvia ad una presentazione specializzata. I risultati vengono organizzati su tre colonne: la denominazione della figura drammatica, la sua descrizione con l'elenco dei personaggi associati e le rispettive opere. Le due modalità di interrogazione del modello analitico del personaggio sono state utilizzate per individuare i casi di legami tra i testi confluiti nel relativo *dataset* (vedi paragrafo 3.2.3.1).

¹⁷⁰ Per facilitare questo tipo di indagine, all'interno della sezione della scheda dell'opera che riporta i personaggi è stata inserita una funzione che permette di aprire la maschera di ricerca per personaggi trovando il campo precompilato con l'elenco dei relativi nomi.

La funzionalità di ricerca dei personaggi non è stata oggetto di valutazione attraverso il questionario in quanto si è ritenuto che, data la sostanziale novità dell'impiego del personaggio come dimensione di analisi, la possibilità da parte degli utenti di coglierne a pieno l'efficacia fosse vincolata ad una formazione specifica sul modello concettuale sottostante ancora ampiamente in fase sperimentale.

5.3 Ricerca avanzata per attributi

La funzione di ricerca libera testuale nasce per fornire uno strumento che semplifichi l'accesso ai contenuti dell'archivio. Se è vero che non viene richiesta all'utente una conoscenza pregressa del modello descrittivo per elaborare una strategia di ricerca efficace, questo per contro limita la possibilità di specificare criteri basati sulla corrispondenza diretta tra attributo e valore ricercato. La funzione di 'ricerca avanzata' è disegnata come strumento complementare alla ricerca libera per consentire all'utente esperto di elaborare criteri di ricerca sulla base di coppie attributo-valore.

Come previsto dal [REQ1], il sistema deve consentire un accesso diretto alle tre sezioni dell'archivio attraverso criteri di ricerca che operino sugli attributi di ciascuna delle tre tipologie di entità principali del dominio. È necessario quindi individuare per ciascuna di esse il sottoinsieme di attributi utilizzabili nella ricerca avanzata. Questa selezione viene implicitamente fatta nelle funzioni di ricerca del sistema tradizionale attraverso l'inclusione o meno di un attributo nella maschera di ricerca. Nel nuovo contesto si rende necessario esplicitarla. Questo per far sì che vengano proposti all'utente solo quegli attributi che sono maggiormente discriminanti per l'individuazione delle istanze di ciascuna tipologia di entità. Come per la rappresentazione delle proprietà in fase di consultazione (vedi paragrafo 2.5.3) si è ricorso ad un livello descrittivo intermedio per definire quali proprietà impiegare come attributi per la composizione di criteri specifici per opere, eventi e documenti.

Nell'ontologia di dominio è stata introdotta una proprietà *ancestor* denominata "CPA1_searchable" (corago:CPA1_searchable) con l'obiettivo di raggruppare sotto di sé le proprietà per cui gli utenti possono indicare in modo puntuale il valore oggetto della ricerca. Ad esempio, se si vuole ricercare il termine "Venezia" come luogo di prima rappresentazione di un'opera, si dovrà individuare il grafo che mette in relazione l'opera (ovvero le entità appartenenti alla classe frbroo:F1_Work) con l'etichetta delle relative località di prima rappresentazione (proprietà rdfs:label delle istanze della classe frbroo:F9_Place)¹⁷¹.

A questo provvede la proprietà corago:CPC77_work_first_performance_place definita come sotto-proprietà (rdfs:subPropertyOf) sia della corago:CPA2_navigable che della corago:CPA1_searchable. La proprietà specifica la *property chain* corrispondente all'attributo che identifica il 'luogo di prima rappresentazione' di un'opera. In questo modo la proprietà assume la doppia funzione di rappresentare sia la forma sintetica del concetto di 'luogo prima rappresentazione', che di proprietà impiegata in fase di ricerca e di visualizzazione del valore dell'attributo. Per l'implementazione delle procedure di interrogazione, in modo speculare a quelle sviluppate per la rappresentazione dei contenuti, il sistema utilizza la forma sintetica della proprietà per la costruzione della maschera di ricerca e la forma completa definita nella sua *property chain* per l'interrogazione SPARQL che viene eseguita.

Gli attributi attraverso cui specificare i criteri analitici per ciascuna delle tre classi principali del dominio è quindi composto dal sottoinsieme delle proprietà definite come 'ricercabili' (ovvero definite come rdfs:subPropertyOf della corago:CPA2_searchable). In questo modo è possibile stabilire a livello di modello

¹⁷¹ Nello specifico il grafo in oggetto risulta articolato in tre archi costituiti dalle proprietà: crm:R16i_was_initiated_by, corago:CP11_first_performance_place, rdfs:label.

di dominio quali attributi possono essere utilizzati come criteri avanzati di selezione delle istanze di ciascuna delle tre classi¹⁷².

Ricerca delle entità in base ai loro attributi

Tipo di entità da ricercare

Titolo

Autore

Personaggio

Descrizione personaggio

Arco temporale della ricerca:

da:

a:

parole in AND

parole in AND

parole in AND

parole in AND

AND

AND

AND

AND

Nota: è possibile concatenare più criteri di ricerca utilizzando gli operatori AND e OR. Per ciascun criterio si può specificare se i singoli termini devono essere presenti distintamente, in alternativa o come frase esatta. La ricerca avviene per parole esatte. È possibile utilizzare come carattere di troncamento "*" (punto+asterisco). Ad esempio inserendo la stringa "giovanni", il sistema presenterà come risultati: "Don Giovanni", "Re Giovanni" o "Giovanni Giscala".

ATTENZIONE: in via sperimentale il sistema limita a 100 i possibili risultati ottenibili dalla ricerca.

Figura 5-6: Maschera per la ricerca avanzata

Dal punto di vista funzionale la maschera per la ricerca ‘avanzata’ (o ‘per attributi’), illustrata nella figura 5.4, è suddivisa in quattro sezioni: la selezione del tipo di entità per cui effettuare la ricerca; la sezione per l’inserimento dei valori degli attributi da utilizzare come criteri; la configurazione del criterio complessivo ottenuto dalla concatenazione di termini e attributi secondo gli operatori logici; l’indicazione dell’arco temporale di riferimento della ricerca. La scelta della tipologia di entità determina i corrispondenti campi per i quali l’utente può indicare il valore desiderato. Il sistema prevede la possibilità di comporre un’interrogazione basata su un massimo di quattro criteri testuali, con contenuti multi-termine concatenabili in AND o OR oppure utilizzabili come frase esatta, a loro volta concatenabili in AND o OR. In aggiunta a questi è previsto la possibilità di specificare il vincolo temporale¹⁷³.

Alcune scelte di natura funzionale, come limitare il numero massimo di attributi concatenabili, nascono da vincoli di natura strettamente tecnica. Il listato 5.2 riporta un esempio di ricerca di opere in base al titolo ed al nome di uno degli autori responsabili. Come si può osservare nel blocco 1, le proprietà che rappresentano i due attributi vengono tradotte nella forma estesa corrispondente alla struttura fisica del grafo. Al di là delle diverse formulazioni possibili per la query, resta la necessità di operare n giunzioni con n pari al numero di attributi indicati. Questo tipo di costruito diventa rapidamente poco scalabile, da ciò deriva la necessità di limitare il numero massimo di attributi utilizzabili come criteri di ricerca¹⁷⁴.

¹⁷² In questo momento, nel Corago SM, sono previste ventinove proprietà utilizzabili per la ricerca avanzata.

¹⁷³ Anche in questo caso viene adottata la logica di restrizione per la dimensione temporale descritta nel paragrafo 5.1.

¹⁷⁴ Sperimentalmente si è verificato che su un grafo dell’ordine dei 40M di triple e l’architettura priva di acceleratori GPU il limite massimo per rientrare in un tempo di esecuzione inferiore al secondo è di quattro criteri distinti.

Template di query SPARQL per la ricerca per attributi

#BLOCCO 0: selezione dei risultati

```
SELECT distinct ?s (rdfs:label as ?p) ?o {
```

#BLOCCO 1: selezione degli attributi su cui effettuare la ricerca

```
{select ?s (group_concat(concat(str(?trg),'&&',str(?subj)) ;separator="###" as ?o)
```

```
{
```

#Vincolo sulla tipologia di entità: Opera

```
{select ?s (REPLACE(str(?subj), ",", "")) as ?subj} ?trg {
```

#Vincolo sul "Titolo dell'opera": proprietà ricercabile corago:CPC20_has_title

```
?s a frbroo:F1_Work.
```

```
?s ecrm:P102_has_title ?trg. ?trg rdfs:label?|rdf:value ?subj}
```

```
}
```

```
UNION
```

```
{select ?s (REPLACE(str(?subj), ",", "")) as ?subj} ?trg {
```

#Vincolo sulla tipologia di entità: Opera

```
?s a frbroo:F1_Work.
```

#Vincolo sul "Autore dell'opera": proprietà ricercabile corago:CPC1_concepted_by

```
?s frbroo:R16i_was_initiated_by/corago:CP2_carried_out_role/corago:CP3_carried_out_actor ?trg. ?trg
```

```
rdfs:label?|rdf:value ?subj}
```

```
}
```

```
} group by ?s}
```

#BLOCCO 2: applicazione criterio di filtro in base ai valori inseriti

```
FILTER ( regex(str(?o),"Otello","i") && regex(str(?o),"verdi","i")).
```

#BLOCCO 3: applicazione criterio di temporale

```
OPTIONAL {?t a <http://erlangen-crm.org/current/E49_Time_Appellation>. ?t rdf:value ?tl. FILTER(?tl >= "1499-12-31T23:10:04.000Z"^^xsd:dateTime && ?tl < "2000-12-30T23:00:00.000Z"^^xsd:dateTime).
```

```
?d ecrm:P78_is_identified_by/ecrm:P78_is_identified_by* ?t.
```

```
?s (corago:CP9_first_performed_in/ecrm:P4_has_time-span) |
```

```
(ecrm:P4_has_time-span) |
```

```
(frbroo:R16i_was_initiated_by/ecrm:P4_has_time-span) |
```

```
(frbroo:CLR6_should_carry/^ecrm:P94_has_created/ecrm:P4_has_time-span) |
```

```
(^ecrm:P67_refers_to) ?d. }
```

```
}
```

Listato 5.2: Esempio di query avanzata per attributi

Occorre inoltre osservare come il criterio di filtro venga applicato utilizzando la funzione REGEX sul nodo *literal* corrispondente al contenuto testuale della proprietà. Questo approccio mima il comportamento della funzione corrispondente funzione LIKE('testo') del linguaggio SQL (Structured Query Language) utilizzata nei sistemi basati su database relazionali come il sistema tradizionale Corago. Questo, se da un lato permette di comparare direttamente il comportamento dei due sistemi, dall'altro non sfrutta le funzionalità di tipo full-text introdotte nel *repository* RDF. Si è verificato empiricamente che sussistono vincoli di natura implementativa all'utilizzo dell'indicizzazione full-text in questo contesto. Il primo vincolo è legato al significativo degrado delle performance che si ottiene sostituendo il costrutto di FILTER per l'applicazione della restrizione col corrispettivo criterio di *match* che utilizza gli specifici predicati. Nel caso della ricerca libera testuale il problema era stato superato attraverso l'introduzione dell'indicizzazione cumulativa dei contenuti testuali (vedi paragrafo 2.6). L'indice testuale però non discrimina il tipo di relazione che intercorre tra testo e risorsa. Questo ne limita l'utilizzo nel caso in cui il criterio di ricerca richiede di operare in base alla corrispondenza attributo-valore. Resta quindi ancora non implementata nel sistema sperimentale la possibilità di effettuare ricerche per attributi che operino in modalità full-text.

D'altro canto, l'obiettivo prevalente della funzione di 'ricerca avanzata' è dimostrare l'equivalenza in termini funzionali delle ricerche 'per attributi' effettuabili su un sistema basato su un modello concettuale a semantica esplicita rispetto al modello relazionale originale da cui deriva. La soluzione proposta inoltre dimostra come la struttura delle proprietà per la navigazione risulti efficace anche come strumento di accesso in fase di interrogazione della base di conoscenza. Questa ambivalenza, formalmente rappresentata attraverso l'ereditarietà multipla, consente all'esperto di dominio di individuare gli attributi chiave per ciascuna tipologia di entità del dominio. Ad esempio, il titolo di un'opera potrà essere definito `corago:CPC20_has_title` sia come proprietà di navigazione identitaria (vedi paragrafo 2.5.3) che come proprietà ricercabile. Si potrà inoltre definire una corrispondente proprietà di indicizzazione full-text (`corago:CPBC20_title`) per inserire il titolo nell'indice testuale. Il risultato è la formalizzazione della valenza del titolo dell'opera sia come elemento informativo che come criterio discriminante in fase di ricerca. La conseguenza è che un qualsiasi sistema che acceda alla base di conoscenza potrà fare riferimento a questo duplice livello descrittivo per ricavare i criteri fondamentali per la sua consultazione.

5.4 Rappresentazione e navigazione multidimensionali

La trasposizione dei contenuti dell'archivio nel formato RDF ha prodotto una organizzazione dell'informazione interamente basata su triple. Ne è derivata una struttura reticolare di rinvii tra nodi informativi rispetto alla quale ciascuno di essi può costituire il punto di ingresso da cui avviare l'esplorazione del grafo (vedi paragrafo 2.5). Data questa libertà di accesso all'informazione, la definizione degli strumenti di navigazione diventa una scelta di natura puramente funzionale che il gestore dell'archivio può liberamente orientare in funzione dei profili di utenza del sistema. Allo stesso tempo, l'ampliamento delle possibilità di navigazione presenta una serie di problematiche che devono essere affrontate per proporre delle strategie di accesso efficaci. Se si vuole consentire all'utente di esplorare liberamente il grafo, il sistema di consultazione dovrà essere in grado di gestire l'eterogeneità dei modelli descrittivi delle diverse entità. Questo comporta la necessità di realizzare tante viste quante sono le classi previste dal modello concettuale sottostante. La soluzione proposta per raggiungere questo grado di flessibilità è far guidare la rappresentazione dei contenuti da parte delle applicazioni di consultazione dalla semantica di dominio (vedi paragrafo 2.5.2). L'efficacia di questo approccio, misurata in termini di comparativi rispetto alla soluzione tradizionale basata su viste specializzate, è stata valutata in base all'apprezzamento da parte degli utenti (vedi quesito Q 2.3 del questionario).

D'altra parte, la libertà di muoversi seguendo percorsi non predeterminati può portare al rischio di 'perdersi tra i contenuti'. Il sistema deve quindi fornire all'utente il 'contesto informativo' necessario per orientarsi. Un esempio in questo senso è dato dall'introduzione dei riferimenti alle occorrenze individuate nelle ricerche testuali. Oltre a questo, però il sistema non pone limiti all'esplorazione a partire dal risultato. È opportuno però valutare quanto sia utile questa flessibilità nella navigazione (vedi quesito 2.4 del questionario) e se non sia necessaria l'individuazione di ulteriori strumenti che la orientino.

Come illustrato nei paragrafi precedenti, gli strumenti di ricerca costituiscono il principale punto di accesso per la consultazione dell'archivio¹⁷⁵. Nel sistema sperimentale sono state introdotte funzioni che impiegano dimensioni trasversali con l'obiettivo di avviare la navigazione a partire da chiavi di accesso che hanno dimostrato una loro significativa efficacia (vedi ad esempio paragrafo 3.2). Si tratta di capire però se queste

¹⁷⁵ I dati relativi agli accessi alle pagine nel periodo dal 1/3/2019 al 31/08/2019 vedono la ricerca per personaggi (`/corago/analyticalSearch`) come la più consultata (1.7k accessi), seguita dalla ricerca libera testuale (`/corago/remoteSearch/corago/remoteSearch`) con 132 accessi e a seguire la ricerca avanzata (34) e la spazio-tempo (33). Nel complesso il sistema è stato utilizzato da oltre 200 utenti distinti (dati ricavati in base a Google Analytics).

costituiscano un effettivo avanzamento rispetto alle modalità di accesso ai contenuti anche dal punto di vista dell'utente finale (vedi quesito 2.5 del questionario).

5.4.1 Gli elenchi di risultati

I risultati prodotti dalle diverse funzioni di ricerca vengono presentati attraverso una maschera comune (figura 5.7). La pagina adotta una struttura tabellare in cui ogni riga rappresenta un risultato della ricerca effettuata. I contenuti sono organizzati in sei colonne distinte: il riferimento alla risorsa, i responsabili, la data di riferimento, la tipologia della risorsa, gli attributi identitari ed i riferimenti al contesto informativo in base al quale il risultato è stato individuato.


Risultato	Responsabili	Data	Tipo	Attributi	Riferimenti
Il "tabarro" 	compositore: Puccini, Giacomo librettista: Adami, Giuseppe	1918	Opera	macrogenere: melodramma genere in fonte: opera prima rappresentazione: Il "trittico" lingua: italiano	"Puccini, Giacomo" (Autore dell'opera) "Il "tabarro" (libretto)
Il "trittico" 	compositore: Puccini, Giacomo	14/12/1918	Evento	luogo: New York sede: Metropolitan Opera	"Puccini, Giacomo" (Autore) "Il "tabarro" (Rappresentazione di)
Il "tabarro - Suor Angelica - Gianni Schicchi" 	compositore: Puccini, Giacomo	1918	Documento	lingua: italiano	"Puccini, Giacomo" (Autore) "Tabarro 1918MI2 - I-SORmde - LC.249.a1" (esemplari) "Il "tabarro - Suor Angelica - Gianni Schicchi. - Milano : G. Ric ..." (Intestazione di sezione)

Figura 5-7: Pagina dei risultati delle ricerche

Nella prima colonna vengono riportate le 'denominazioni' delle risorse (proprietà `rdfs:label`) con il relativo link che rinvia alla scheda di dettaglio. L'icona sottostante rinvia invece alla relativa scheda del sistema tradizionale per dare la possibilità di confrontare le due versioni. La seconda colonna riporta gli autori ed il relativo ruolo; la terza la datazione di riferimento. Nella quarta viene indicata a quale delle tre tipologie di entità previste il risultato appartiene. La colonna successiva presenta gli attributi utili a identificare l'opera (gli 'attributi identitari', vedi paragrafo 2.5.3). Nell'ultima colonna vengono indicati i contenuti testuali corrispondenti ai criteri impostati per la ricerca. Ciascuno dei riferimenti rinvia alla relativa risorsa (vedi la modalità di indicizzazione testuale delle risorse descritta nel paragrafo 2.6) permettendo all'utente di verificare e valutare i contenuti individuati e la rilevanza rispetto agli obiettivi della ricerca.

Le prime tre colonne prevedono la possibilità di ordinare i risultati in base ai rispettivi contenuti. Le prime due in base all'ordinamento alfabetico, la terza in base alla data. I risultati sono suddivisi in pagine di 10 elementi, lo scorrimento e la navigazione tra le diverse pagine è reso possibile da una barra di controllo alla base della maschera.

L'impostazione standard per la navigazione dei link presenti nella pagina prevede l'apertura di nuove schede del browser per ciascun risultato visitato. Lo scopo è consentire all'utente di utilizzare la pagina dei risultati come punto di partenza per l'approfondimento dei risultati attraverso la consultazione in parallelo delle relative schede (un modello di navigazione che potremmo definire orizzontale o estensivo)¹⁷⁶. L'efficacia dei contenuti ripotati nella pagina e della modalità di navigazione è stata sondata attraverso i quesiti Q 2.1 e Q 2.2 del questionario di valutazione.

Un'ultima funzionalità presente sulla pagina è il lancio della modalità di rappresentazione dei risultati su cartografia. Nell'angolo superiore destro della pagina è presente una icona che permette di aprire il visore spazio-tempo (vedi paragrafo 5.4.3). Attraverso questa funzione il sistema rappresenta l'insieme dei risultati della ricerca sulla cartografia, adottando le coordinate di riferimento associate ai risultati per la loro collocazione e la data di riferimento per ordinarli lungo l'asse temporale. Si passa quindi da una rappresentazione monodimensionale (secondo i criteri di ordinamento previsti dalla tabella) ad una organizzazione bidimensionale basata sulla collocazione spazio-temporale dei risultati della ricerca.

5.4.2 Le schede delle entità

Come discusso nel paragrafo 2.5, tradurre l'informazione contenuta nel grafo RDF in una rappresentazione efficace per l'utente finale non è immediato. La soluzione individuata prevede l'introduzione di uno strato intermedio tra il livello del *dataset* e le viste che ne presentano i contenuti all'utente. La mediazione operata dallo strato intermedio ha come obiettivo la specializzazione della semantica del modello 'generalista' definito attraverso le ontologie di riferimento in base alle caratteristiche specifiche del dominio dell'archivio. Utilizzando le 'proprietà di navigazione', è possibile modellare un punto di vista che produca una prospettiva efficace per l'utente. Il sistema dovrà semplicemente basarsi su quest'ultimo per costruire delle rappresentazioni coerenti delle diverse entità del dominio.

La pagina del sistema sperimentale che produce le schede delle diverse entità implementa l'architettura descritta nel paragrafo 2.5.4. L'ordine di presentazione dei contenuti e la morfologia del *layout* della pagina sono interamente guidati dalla semantica dei contenuti che vengono caricati. Per ciascuna istanza di risorsa il sistema organizza la pagina interpretando le annotazioni delle proprietà di navigazione dei contenuti che vengono mostrati.

L'impostazione della struttura della scheda comune a tutte le tipologie di entità prevede l'articolazione della pagina in una intestazione e in un corpo. L'intestazione riporta la denominazione della risorsa, l'URI e la tipologia della risorsa che viene rappresentata. Nel corpo, suddivise per area tematica, vengono mostrate le proprietà valorizzate per l'istanza corrente. Il *layout* e i contenuti della scheda di ciascuna risorsa sono quindi strutturalmente correlati alla tipologia ed ai contenuti effettivamente presenti del suo grafo. La concorrenza di questi due fattori, se da una parte rende il sistema in grado di costruire viste in funzione della morfologia dei contenuti in modo dinamico, dall'altra produce una significativa variabilità del peso computazionale necessario a generare la corrispondente rappresentazione. Per mantenere i tempi di risposta entro limiti compatibili con la fruizione da parte dell'utente, il sistema gestisce il caricamento dei contenuti della scheda in modo incrementale¹⁷⁷. La scheda viene quindi composta a partire dall'intestazione e a seguire vengono caricati i contenuti delle singole aree tematiche in modalità asincrona. L'effetto percepito dagli utenti è il comporsi progressivo della scheda via via che i dati vengono caricati e visualizzati.

¹⁷⁶ Si osservi come invece nella navigazione dei riferimenti presenti sulla scheda della risorsa il sistema preveda, come impostazione di base, di aprire la nuova scheda nello stesso pannello (*tab*) del browser (in questo caso seguendo un modello verticale).

¹⁷⁷ Il caricamento incrementale è ottenuto attraverso una interazione asincrona tra browser e server basata sulla tecnologia JavaScript and XML (Ajax). Lato server è stato esteso il *controller* già previsto dalla soluzione LodView (vedi appendice B.1) per includere i metodi necessari a fornire i sottoinsiemi di dati richiesti dalle funzioni di visualizzazione.

Otello
<http://corago.unibo.it/resource/BRANI/Z000082306>

entità di tipo: F15_Complex_Work

Dati di base
 macrogenere: melodramma
 genere in fonte: dramma lirico
 concepito da: **compositore:** Verdi, Giuseppe (data di nascita: 1813/10/10; data di morte: 1901/01/27)
librettista: Boito, Arrigo (data di nascita: 1842/02/24; data di morte: 1918/06/10)
 prima rappresentazione: Otello (data: 1887/02/05; luogo: Milano; sede: Teatro alla Scala; produzione: prima assoluta; autore dell'opera rappresentata: Verdi, Giuseppe)
 titolo: Otello
 lingua: italiano
 tipo testo letterario: nuovo
 ambientazione: Scena: Una città di mare nell'isola di Cipro. Epoca: La fine del secolo XV.
 collocazione temporale: 1380

Personaggi

N.	Personaggio	Qualifica personaggio
1	Otello	moro, generale dell'Armata Veneta
2	Jago	alfiere
3	Cassio	capo di squadra
4	Roderigo	gentiluomo veneziano
5	Lodovico	ambasciatore della Repubblica Veneta
6	Montano	predecessore d'Otello nel governo dell'isola di Cipro
7	un araldo	
8	Desdemona	moglie d'Otello
9	Emilia	moglie di Jago

cori e comparse:

Soldati e marinai della Repubblica veneta, gentildonne e gentiluomini veneziani, popolani ciprioti d'ambo i sessi; uomini d'arme greci, dalmati, albanesi, fanciulli dell'isola, un taverniere. Quattro servi di taverna,

ricerca opere correlate

▶ QUADRI SCENICI E STRUTTURA

▶ RAPPRESENTAZIONI

▶ DOCUMENTI

▶ RELAZIONI

Figura 5-8: Esempio di scheda di un'opera

La figura 5.8 illustra la scheda relativa all’*Otello* di Giuseppe Verdi. Sotto l’intestazione, vengono riportate le prime due ‘aree tematiche’ della scheda dell’opera (ovvero, in questo caso, una entità del tipo frbroo:F15_Complex_Work come riportato nell’intestazione in alto a destra): i “dati di base” e i “personaggi”. Sotto le prime due aree, compaiono i pannelli espandibili relativi a: “quadri scenici e struttura”, “rappresentazioni”, “documenti” e “relazioni”. L’utilizzo di pannelli espandibili è dettato dalla necessità di rendere la prima visualizzazione il più compatta possibile: l’utente sceglierà quale sezione espandere in funzione dei suoi interessi.

La differenza nella modalità di rappresentazione delle diverse aree tematiche è gestita mediante *widget* dedicate (vedi paragrafo 2.5.4). Nel caso illustrato, l’area “dati di base” prevede una semplice elencazione delle proprietà di navigazione afferenti secondo il rispettivo ordine di rilevanza (ad esempio il macro-genere: corago:CPC12_work_genre). L’elenco dei personaggi, definito in base alla proprietà corago:CPC4_has_character, è strutturato invece in una tabella che riporta nelle diverse colonne: l’ordinale (proprietà corago:CPC42_character_priority), il nome (proprietà rdfs:label) e la descrizione (proprietà corago:CPC31_has_note).

Per la visualizzazione della struttura dell’opera è stata utilizzata una diversa rappresentazione implementata nel pannello “quadri scenici e struttura” (vedi figura 5.9). La singola sezione dell’opera, definita dalla proprietà corago:CPC28_has_part, viene rappresentata nella struttura gerarchica in base al relativo ordinale (proprietà corago:CPC58_part_number). Ogni sezione è corredata dal relativo tipo (proprietà corago:CPC55_fragment_type) e dalla nota descrittiva.

▾ QUADRI SCENICI E STRUTTURA

sezione:

1 - Atto primo (Atto)

1.1 - L'*esterno del castello (Quadro scenico)

L'esterno del castello. Una taverna con pergolato. Gli spaldi nel fondo e il mare. È sera. Lampi, tuoni, uragano. \$ \$ \$ Scena I - dalla scala della spiaggia - da un lato una catasta di legna - dalla catasta incominciano ad alzarsi dei globi di fumo - il fuoco divampa - tavole della taverna - appenderanno al pergolato dell'osteria delle lanterne veneziane a vari colori che illumineranno gaiamente la scena - il fuoco si spegne a poco a poco - campane a stormo Scena II - fiaccole - le nubi si diradano Scena III - plaga del cielo stellato 11.1

2 - Atto secondo (Atto)

2.1 - Una *sala terrena nel castello (Quadro scenico)

Una sala terrena nel castello. Due vasti veroni ai lati: una porta nel mezzo che dà sul giardino. \$ \$ \$ Scena II - alberi - uscio del lato destro Scena III - vasta apertura del fondo 12.1

3 - Atto terzo (Atto)

3.1 - La *gran sala del castello (Quadro scenico)

La gran sala del castello. A destra un vasto peristilio a colonne. Questo peristilio è annesso ad una sala di minori proporzioni; nel fondo della sala un verone. \$ \$ \$ Scena II - porta di sinistra Scena VIII - accasciato su d'un sedile 13.1

4 - Atto quarto (Atto)

4.1 - La *camera da letto di Desdemona (Quadro scenico)

La camera da letto di Desdemona. Letto, inginocchiatoio, tavolo, specchio, sedie. Una lampada arde appesa davanti all'immagine della Madonna che sta al di sopra dell'inginocchiatoio. Porta a destra. È notte. Un lume acceso sul tavolo. \$ \$ \$ Scena III - porta segreta - scimitarra sul tavolo - face [=fuoco], accesa/spenta - s'ode bussare alla porta 14.1

Figura 5-9: Pannello relativo alla struttura di un'opera

Per la visualizzazione delle rappresentazioni dell'opera si è utilizzata una *widget* composta da due sezioni interdipendenti che presentano gli spettacoli in forma tabellare e su una linea temporale (vedi figura 5.10).

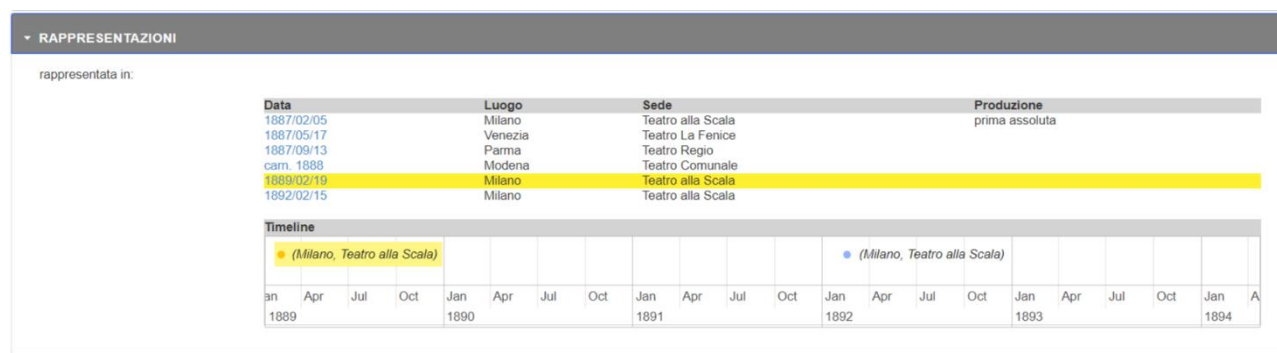


Figura 5-10: Pannello relativo alle rappresentazioni di un'opera

Le due modalità di rappresentazione sono correlate: selezionando un elemento della *timeline* si evidenzia il corrispondente record della tabella e viceversa, l'obiettivo è fornire una doppia prospettiva degli eventi: sinottica e sequenziale.

▼ RAPPRESENTAZIONI			
programma:			
	Num. progressivo	Opera	Responsabili
	1	Otello	Verdi, Giuseppe
	Interpreti		
	Personaggi	Interpreti	Ruolo Vocale
	Otello	Tamagno, Francesco	tenore
	Jago	Maurel, Victor	baritono
	Cassio	Paroli, Giovanni	tenore
	Roderigo	Fornari, Vincenzo	tenore
	Lodovico	Navarini, Francesco	basso
	Montano	Limonta, Napoleone	basso
	un araldo	Lagomarsino, Angelo	basso
	Desdemona	Pantaleoni, Romilda	soprano
	Emilia	Petrovich, Ginevra	mezzosoprano

Figura 5-11: Pannello relativo al programma dell'evento

Nella stessa area tematica, per gli eventi, è riportato il programma (rappresentato in base alla proprietà corago: CPC56_performance_program). In questo caso la *widget*, illustrata in figura 5.11, presenta l'elenco delle opere rappresentate per ciascuna delle quali è possibile visualizzare, in un pannello contestuale, i personaggi con i rispettivi interpreti.

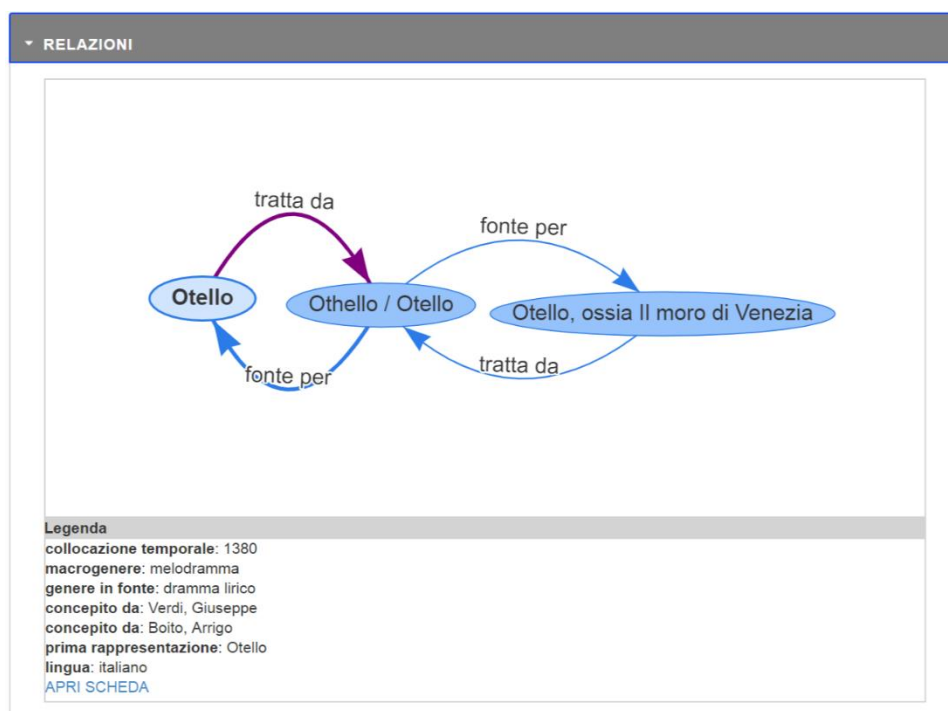


Figura 5-12: Pannello relativo alle relazioni tra le opere

La figura 5.12 raffigura il pannello utilizzato per rappresentare le relazioni tra le opere identificate dalle due proprietà corago: CPC82_work_ancestors e corago: CPC83_work_descendants. La metafora visuale utilizzata in questo caso è direttamente quella del grafo. Il sistema mostra l'opera ed i suoi immediati ascendenti sotto forma di nodi. Selezionando un nodo, in legenda vengono mostrate le relative *features* ed i riferimenti

accedere alla scheda di dettaglio. È possibile espandere il grafo facendo un doppio click su ciascuna opera: in questo modo vengono caricati i nodi relativi alle opere di cui è indicata come fonte¹⁷⁸.

In base a quanto illustrato finora risulta evidente come il sistema di rappresentazione dei contenuti è in grado di gestire il polimorfismo intrinseco delle risorse. L'introduzione di uno strato espressamente dedicato alla modellazione della semantica di rappresentazione dei contenuti è lo strumento con cui costruire dinamicamente viste specializzate in funzione del modello specificato dall'ontologia di dominio. L'applicazione per la consultazione può quindi specializzare le metafore di visualizzazione associandole direttamente alla natura dei rispettivi contenuti facendo unicamente riferimento alle relative 'proprietà di navigazione'.

La rappresentazione risulta quindi associata alla semantica dei contenuti e non direttamente alla loro morfologia. Questo è un significativo scarto rispetto alle soluzioni basate su sistemi relazionali che necessariamente devono basarsi su viste strutturate funzione dell'organizzazione fisica dei dati. L'efficacia di questa soluzione è stata valutata attraverso il quesito Q 2.3 del questionario.

¹⁷⁸ Nel caso in oggetto, il grafo evidenzia come l'*Othello* di Shakespeare sia indicato contemporaneamente fonte per l'*Otello* di Verdi che di quello di Rossini.

5.4.3 Rappresentazione su cartografia

Uno degli ambiti previsti per la sperimentazione ha avuto come oggetto l'impiego di metafore visuali in grado di produrre nuove modalità di rappresentazione dei contenuti dell'archivio. Le attività di normalizzazione delle dimensioni tempo e spazio, assieme alla georeferenziazione dei toponimi (vedi paragrafi 3.1 e 4.3.1), hanno introdotto la possibilità di posizionare le risorse della base di conoscenza sulla cartografia. Per offrire all'utente una visione coordinata delle due dimensioni (come richiesto dal [REQ6]) è stato realizzato un visore cartografico integrato ad una linea temporale¹⁷⁹. Il risultato è una rappresentazione che compone le due prospettive: la distribuzione geografica delle entità e la loro collocazione rispetto al tempo.

L'interfaccia risultante che viene proposta agli utenti è illustrata in figura 5.13.

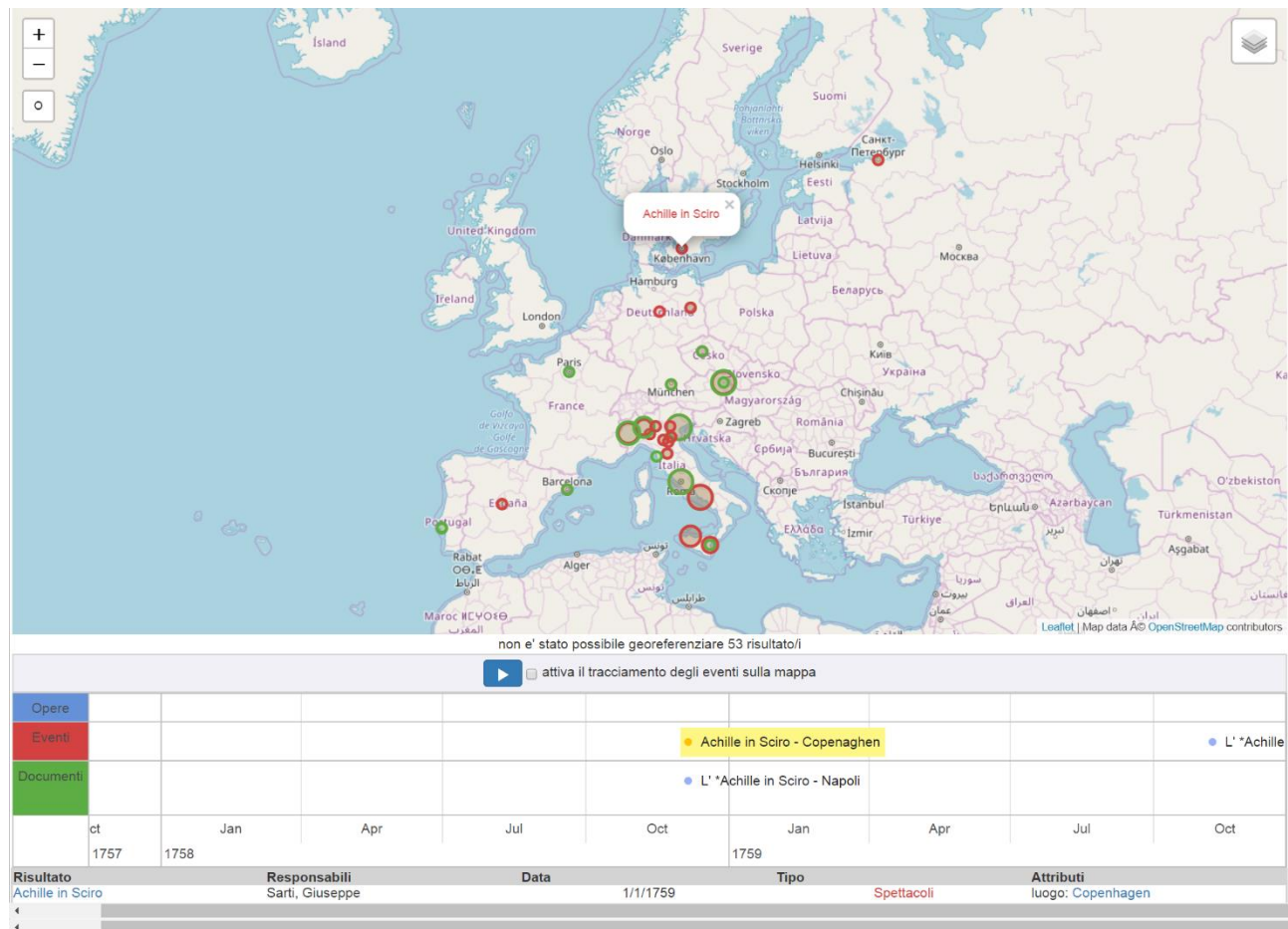


Figura 5-13: Visore spazio-tempo

Sulla cartografia vengono indicate, mediante dei marcatori circolari, la collocazione delle singole risorse. Ciascuna tipologia è rappresentata attraverso un colore: blu per le opere, rosso per gli eventi e verde per i documenti. Il diametro del marcatore è proporzionale al numero di risorse posizionate sulla località (passando il mouse sul marcatore il sistema indica il numero totale di entità). Selezionando il singolo marcatore, il sistema presenta nella tabella a fondo pagina l'elenco analitiche delle entità corrispondenti e le evidenzia in giallo sulla linea temporale.

¹⁷⁹ Il visore cartografico è stato realizzato come componente lato client utilizzando cartografia OpenStreetMap (Haklay e Weber, 2008) e la libreria Leaflet JS (sito web del progetto: <https://leafletjs.com/>, visitato il 29/06/2019). Il visore è implementato come pagina JSP.

Subito sotto la mappa, l'interfaccia propone la vista delle risorse rispetto alla linea temporale. Queste sono collocate su tre fasce corrispondenti alle tre tipologie e posizionate in corrispondenza della data di riferimento (composizione o prima rappresentazione per le opere, rappresentazione per gli eventi ed edizione per i libretti). La linea prevede una risoluzione minima del mese ed una massima di circa un decennio; è inoltre possibile passare in continuità dall'una all'altra effettuando uno zoom. Ogni risorsa è rappresentata da un punto ed una etichetta che indica il titolo e la località. Selezionando la risorsa sulla linea temporale, questa viene evidenziata anche sulla mappa attraverso un *balloon* posizionato sulla località relativa. Selezionando il *balloon* si può accedere direttamente alla scheda di dettaglio.

Sotto la *timeline* viene presentata la tabella che riporta le informazioni relative alle entità selezionate attraverso gli altri due controlli: mappa e linea temporale. Questa è strutturata su quattro colonne che riportano rispettivamente: denominazione, responsabili, data, tipo e gli attributi identitari della risorsa.

La maschera nel suo insieme include delle funzionalità utili a gestire la visualizzazione o a filtrare il numero di risultati. Sulla mappa è presente nell'angolo superiore destro un controllo attraverso cui l'utente può selezionare quali strati (o *layers*) visualizzare. Questi sono, oltre ai tre relativi alle tipologie di entità, il *layer* relativo ai nomi delle località attraverso cui è possibile aprire la corrispondente pagina di DBPedia e i *layer* su cui il sistema traccia sulla mappa le 'rotte' relative all'evoluzione temporale degli eventi (vedi paragrafo 5.4.3.1).

Nell'angolo superiore sinistro, sotto i controlli per la gestione dello zoom, è riportato un pulsante che attiva la possibilità di disegnare sulla mappa un'area circolare di raggio arbitrario. Questo permette all'utente di delimitare l'area rispetto a cui filtrare le entità visualizzate. In questo modo è possibile focalizzare ulteriormente l'analisi su un sottoinsieme di risorse in funzione della loro collocazione. I due strumenti di selezione sono illustrati dalla figura 5.14.

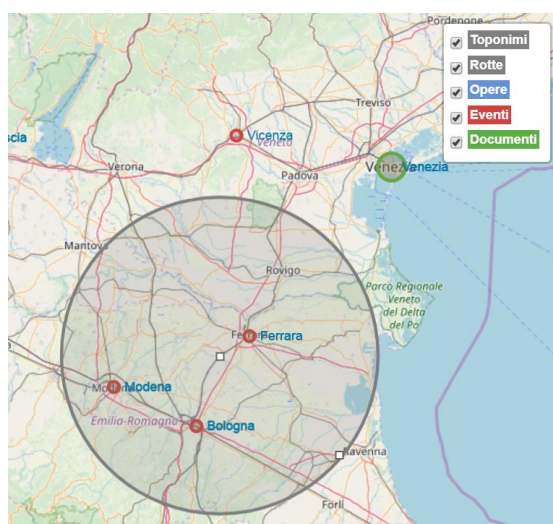


Figura 5-14: Strumenti di selezione del visore cartografico

5.4.3.1 Funzione di tracciamento

L'obiettivo di una rappresentazione integrata delle dimensioni spazio e tempo è riuscire a restituire la dinamica dei fatti rappresentati. Ad esempio, si consideri la carriera di un interprete. I dati presenti in archivio consentono di rappresentare sulla mappa tutte le località degli spettacoli cui ha preso parte. La linea temporale è in grado di articolare la sequenza delle recite nell'arco della carriera. Se la si scorre in modo

lineare seguendo il trascorre del tempo si è in grado di leggere sulla mappa gli spostamenti dell'artista. A questo scopo è stata introdotta una funzionalità che permette di avviare lo scorrimento sequenziale degli eventi sulla *timeline* e, opzionalmente, di segnare sulla mappa gli spostamenti da una località alla successiva. Il risultato è il progressivo delinearsi sulla mappa del peregrinaggio professionale dell'interprete nell'arco della sua attività. La stessa modalità di 'tracciamento' spazio-temporale può essere utilizzata per studiare gli spostamenti di una compagnia di interpreti o la sequenza delle rappresentazioni di una particolare opera.

Di per sé l'adozione di un visore cartografico coordinato ad una *timeline* non rappresenta una novità nell'ambito della visualizzazione dei dati storiografici. La piena integrazione con le funzionalità di ricerca dell'archivio e gli strumenti di rappresentazione dinamica del tempo restituiscono una visione plastica ed innovativa delle vicende del melodramma, di grande efficacia per la rappresentazione di fenomeni storici.

5.5 Valutazione del sistema

L'efficacia da un punto di vista funzionale dell'adozione delle tecnologie semantiche è stata una delle componenti essenziali della sperimentazione svolta.

Nel paragrafo 1.5 sono stati illustrati i requisiti per l'allestimento dell'applicazione di consultazione dell'archivio *Corago* una volta che i suoi contenuti sono stati trasposti in un modello concettuale a semantica esplicita. I requisiti sono stati raccolti in parte come elementi necessari per operare un confronto diretto della soluzione sperimentale rispetto all'approccio tradizionale ed in parte raccogliendo le richieste di evoluzione del sistema tradizionale espresse dagli utenti. La valutazione del raggiungimento dell'insieme di requisiti funzionali e non, come l'usabilità e le performance (vedi paragrafo 1.5.5), è stata effettuata utilizzando un questionario sottoposto ad un campione dell'utenza tipo del sistema tradizionale articolato in profili diversi¹⁸⁰.

Sono stati individuati quattro profili utente:

1. gli **esperti del dominio**: in questa categoria rientrano tutti gli studiosi (accademici e non) del melodramma, i professionisti e gli operatori che abbiano una approfondita conoscenza del dominio. È una utenza che impiega l'archivio *Corago* come strumento di supporto alle proprie attività professionali ed è in grado di contribuire attivamente all'estensione dei contenuti e delle funzionalità del catalogo;
2. gli **archivisti/bibliotecari**: sono utenti professionali, o in formazione, con specifiche competenze biblioteconomiche e archivistiche. L'interesse nei contenuti dell'archivio è prevalentemente concentrato sui documenti e la relativa collocazione tra repertorio e cronologia, pur tangente alle rispettive competenze e attività, costituisce un contenuto informativo aggiuntivo al tradizionale livello descrittivo cui hanno accesso;
3. il **'melomane'**: è l'utente che consulta l'archivio per proprio arricchimento personale. Utente informato e pratico del dominio, ricerca nell'archivio la possibilità di esplorarne i contenuti seguendo il filo dei propri interessi;
4. l'**utente generico**: figura non necessariamente avvezzata ai contenuti, tipicamente imbattutasi nel sistema attraverso ricerche su motori generalisti, sperimenta le funzionalità del sistema per curiosità o approfondimento di quanto già individuato su altre fonti.

Sulla base delle definizioni dei profili è stato costituito il *panel* a cui è stato somministrato il questionario con la seguente composizione:

Tabella 5.1: Numero soggetti per profilo

Profilo	Numero
Esperti di dominio (ESP)	2
Archivisti/bibliotecari (ARC)	2
Melomani (MEL)	1
Utenti generici (GEN)	3
TOTALE	8

¹⁸⁰ Nell'allestimento del questionario, del panel di utenti e nell'analisi dei risultati si è fatto riferimento alla metodologia di rilevamento quantitativo applicata in sociologia. Come testo di riferimento è stato adottato il manuale di P. Corbetta, *Metodologia e tecniche della ricerca sociale*, Il Mulino, 2014.

Una volta definito il campione ed individuati i soggetti per ciascun profilo, il rilevamento è avvenuto in forma anonima.

5.5.1 Struttura del questionario di valutazione

Il questionario è stato impostato in termini comparativi con l'obiettivo di misurare il gradimento del nuovo sistema "Corago LOD" rispetto all'applicazione tradizionale di consultazione.

Dove la copertura funzionale tra il sistema tradizionale e il nuovo sono sovrapponibili il questionario vuole rilevare il gradimento in termini comparativi. Per le nuove funzionalità introdotte come: le ricerche spazio-tempo, la ricerca per personaggi e gli strumenti di visualizzazione su cartografia, il gradimento viene rilevato in termini assoluti in assenza di un termine di paragone.

Il questionario è articolato in tre sezioni. La prima serie di domande si concentra sulla rilevanza dei risultati ottenuti a fronte di una ricerca, mentre la seconda sulla fruibilità percepita delle modalità di presentazione e navigazione dei dati presentati. La terza parte consiste in una domanda aperta volta a raccogliere suggerimenti su aspetti ad oggi mancanti o, comunque, suscettibili di miglioramento rispetto al sistema sperimentale.

5.5.1.1 Sezione prima: ricerca dei contenuti

La prima sezione si compone di cinque quesiti chiusi con valutazione espressa attraverso termini qualitativi ricondotti a valori numerici scalari tra 1 e 5.

Quesito 1.1: "Partendo dallo stesso insieme di criteri di ricerca, quanto i risultati proposti da ciascuno dei due sistemi risultano coerenti con i criteri immessi? Marca con una X la casella corrispondente al tuo giudizio per ciascuno dei due sistemi."

Valori da assegnare a ciascuno dei due sistemi: "Affatto" = 1, "Poco" = 2, "Abbastanza" = 3, "Molto" = 4, "Completamente" = 5.

Obiettivo: misurare la *precision* percepita dall'utente in termini comparativi tra i due sistemi.

Quesito 1.2: "Partendo dallo stesso insieme di criteri di ricerca, quanto i due sistemi sono risultati efficaci nell'individuare risultati per te utili? Marca con una X la casella corrispondente al tuo giudizio per ciascuno dei due sistemi."

Valori da assegnare a ciascuno dei due sistemi: "Affatto" = 1, "Poco" = 2, "Abbastanza" = 3, "Molto" = 4, "Completamente" = 5.

Obiettivo: misurare la *recall* percepita dall'utente in termini comparativi tra i due sistemi.

Quesito 1.3: "Nell'effettuare le ricerche sui due sistemi, come giudicheresti le performance in termini di rapidità di risposta dei due sistemi? Marca con una X la casella corrispondente al tuo giudizio per ciascuno dei due sistemi."

Valori da assegnare a ciascuno dei due sistemi: "Scadenti" = 1, "Mediocri" = 2, "Sufficienti" = 3, "Buone" = 4, "Ottime" = 5.

Obiettivo: misurare la percezione della performance da parte degli utenti in termini assoluti.

Quesito 1.4: “Nel sistema Corago LOD, quanto è risultato facile raffinare le ricerche utilizzando criteri più restrittivi per ridurre il numero di risultati prodotti? Marca con una X la casella corrispondente al tuo giudizio.”

Valori da assegnare a Corago LOD: “Molto difficile” = 1, “Difficile” = 2, “Abbastanza facile” = 3, “Facile” = 4, “Molto facile” = 5.

Obiettivo: misurare la facilità con cui l’utente è in grado di ottenere un miglioramento della *precision* delle ricerche effettuate col nuovo sistema.

Quesito 1.5: “Nel sistema Corago LOD, quanto è risultata utile la funzionalità di ricerca spazio-tempo? Marca con una X la casella corrispondente al tuo giudizio.”

Valori da assegnare a Corago LOD: “Molto difficile” = 1, “Difficile” = 2, “Abbastanza facile” = 3, “Facile” = 4, “Molto facile” = 5.

Obiettivo: misurare in modo assoluto la l’utilità percepita della nuova modalità di ricerca spazio-tempo.

5.5.1.2 Sezione seconda: navigazione dei contenuti

La seconda sezione si compone di cinque quesiti chiusi con valutazione espressa attraverso termini qualitativi ricondotti a valori numerici scalari tra 1 e 5.

Quesito 2.1: “Quanto la rappresentazione delle informazioni che i due sistemi presentano nelle rispettive pagine dei risultati delle ricerche risulta leggibile e chiara? Marca con una X la casella corrispondente al tuo giudizio per ciascuno dei due sistemi.”

Valori da assegnare a ciascuno dei due sistemi: “Affatto” = 1, “Poco” = 2, “Abbastanza” = 3, “Molto” = 4, “Completamente” = 5.

Obiettivo: misurare in modo comparativo la leggibilità dei risultati. Atteso un vantaggio derivante dall’indicizzazione semantica dei risultati da rilevare attraverso una correlazione con la successiva domanda 2.2.

Quesito 2.2: “Nella ricerca a testo libero del sistema Corago LOD, quanto risultano utili i riferimenti alle relazioni tra le corrispondenze testuali individuate ed i risultati presentati? Marca con una X la casella corrispondente al tuo giudizio.”

Valori da assegnare a Corago LOD: “Affatto” = 1, “Poco” = 2, “Abbastanza” = 3, “Molto” = 4, “Fondamentali” = 5.

Obiettivo: misurare in modo assoluto l’utilità percepita del ‘contesto informativo’ ottenuto in base ai ‘full text breadcrumbs’.

Quesito 2.3: “Quanto risultano leggibili le schede delle entità Opera, Evento e Documento proposte da ciascuno dei due sistemi? Marca con una X la casella corrispondente al tuo giudizio per ciascuno dei due sistemi.”

Valori da assegnare a ciascuno dei due sistemi: “Affatto” = 1, “Poco” = 2, “Abbastanza” = 3, “Molto” = 4, “Perfettamente” = 5.

Obiettivo: misurare in modo comparativo la leggibilità dei risultati nella forma proposta dai due sistemi.

Quesito 2.4: “Nel sistema Corago LOD, quanto risulta utile la possibilità di navigare tra le diverse schede delle entità? Marca con una X la casella corrispondente al tuo giudizio.”

Valori da assegnare a Corago LOD: “Affatto” = 1, “Poco” = 2, “Abbastanza” = 3, “Molto” = 4, “Fondamentale” = 5.

Obiettivo: misurare in modo assoluto l'utilità percepita della possibilità di navigare il grafo sottostante ciascun risultato.

Quesito 2.5: “Nel sistema Corago LOD, quanto risultano le funzionalità per la rappresentazione su cartina dei risultati delle ricerche? Marca con una X la casella corrispondente al tuo giudizio.”

Valori da assegnare a Corago LOD: “Affatto” = 1, “Poco” = 2, “Abbastanza” = 3, “Molto” = 4, “Fondamentali” = 5.

Obiettivo: misurare l'utilità percepita della rappresentazione dei dati sulla cartografia.

5.5.1.3 Sezione terza: suggerimenti dell'utente

La terza sezione è costituita da una unica domanda a risposta aperta, non obbligatoria, che recita: “In continuità ai giudizi espressi fino ad ora, secondo te, come potrebbe migliorare il sistema 'Corago LOD'?”.

5.5.2 Modalità di somministrazione e rilevamento dei risultati

In funzione della diversa competenza di ciascuno dei profili, prima della somministrazione si è provveduto a fornire una introduzione ai contenuti dell'archivio e a contestualizzare i due sistemi oggetto della valutazione.

I contenuti dell'introduzione ed il livello di approfondimento sono stati tarati in funzione del profilo: si è partiti dalla semplice presentazione del questionario agli utenti esperti, ad una breve introduzione al modello concettuale in base a cui sono strutturati i contenuti dei due archivi per gli utenti generici.

A valle della fase introduttiva tenuta dall'autore, di norma in presenza, ad un gruppo di soggetti appartenenti allo stesso profilo, la somministrazione è avvenuta in forma anonima. La tracciabilità del questionario è quindi esclusivamente a livello di profilo.

Agli utenti è quindi stato dato modo di accedere al sistema sperimentale ed al sistema tradizionale con strumenti propri essendo entrambi liberamente accessibili via internet. In particolare, per rispondere ai quesiti proposti, è stato suggerito agli utenti di individuare dei casi di studio relativi a contenuti affini a propri interessi ed utilizzarli per sondare l'efficacia nel reperimento di contenuti. L'interazione degli utenti con i due sistemi è stata quindi prevalentemente incentrata su attività di ricerca e consultazione dei risultati attraverso le interfacce grafiche messe a disposizione. Non sono stati posti limiti temporali alla sperimentazione, né la stessa è stata condotta in modo supervisionato né in ambiente controllato. È stata inoltre lasciata la possibilità di effettuare la restituzione del questionario a valle di ulteriori attività di sperimentazione da condursi in autonomia in separata sede. L'obiettivo è stato cercare di avvicinare l'esperienza dello sperimentatore a quella dell'impiego degli strumenti sottoposti a valutazione nel contesto delle proprie attività di studio o di fruizione libera dell'archivio.

5.5.3 Analisi quantitativa dei risultati

In questa sezione vengono presentati i risultati ricavati dalla somministrazione di 8 questionari, suddivisi per profilo (vedi dettaglio riportato nella tabella 5.1).

I risultati sono riportati in forma di matrice i punteggi di ciascun quesito aggregati per profilo. Si è ottenuta una matrice di Profilo/Punteggio da cui è stata ricavata la misura relativa calcolata come media del punteggio ottenuto per ciascun profilo. Dalle medie per profilo è stata a sua volta calcolata la media generale del punteggio del quesito. Nel caso di quesiti comparativi le due matrici sono suddivise per ciascuno dei due sistemi oggetto della valutazione.

Dalle misure relative per profilo/sistema è stato ricavato un istogramma per quesito e dalla media generale dei quesiti è stato ricavato un istogramma di sintesi del questionario (diagramma 5.1).

Sulla base di questa organizzazione dei rilevamenti si è proceduto all'analisi monovariata¹⁸¹ dei singoli quesiti.

I questionari compilati dagli utenti e il foglio di calcolo utilizzato per l'analisi sono pubblicati all'indirizzo: <https://doi.org/10.6084/m9.figshare.9466505>.

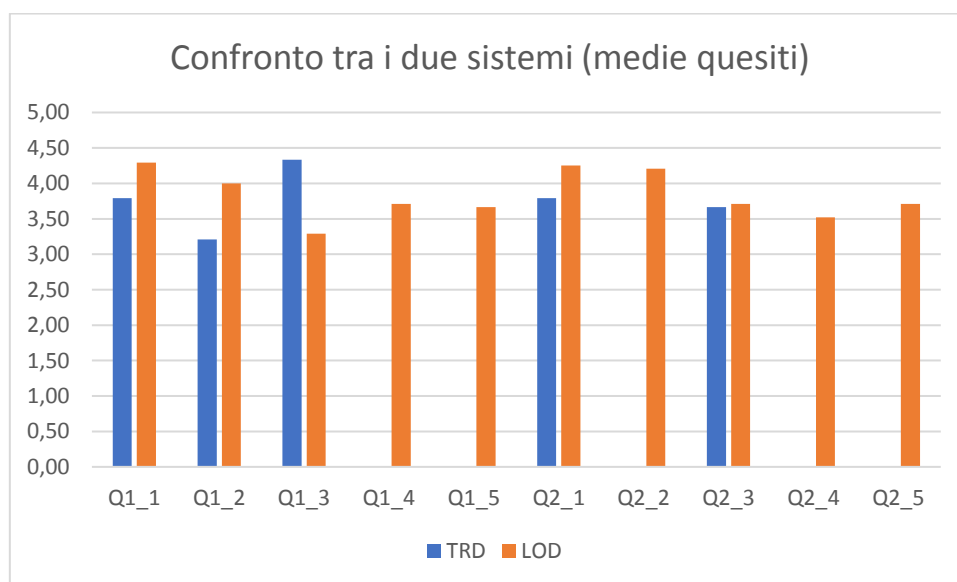


Diagramma 5.1: Quadro di sintesi dei risultati del questionario di valutazione

¹⁸¹ "L'analisi monovariata è un'analisi puramente descrittiva dei fenomeni studiati, che si limita ad esporre come ogni variabile è distribuita fra i casi rilevati, senza porsi problemi sulle relazioni tra le variabili. [...] Essa inoltre rappresenta una prima descrizione dei fenomeni analizzati e contribuisce alla comprensione della struttura del campione e della sua rappresentatività" (Corbetta, 2014).

5.5.3.1 Efficacia nella ricerca dei contenuti

5.5.3.1.1 Q 1.1 – Coerenza dei risultati coi criteri immessi

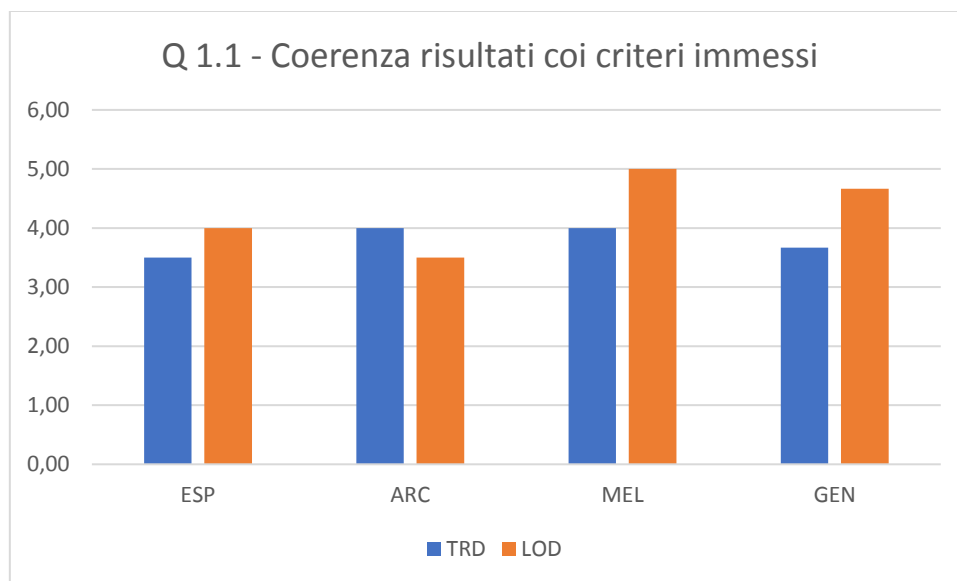


Diagramma 5.2: Q 1.1 - Coerenza risultati coi criteri immessi

Il quesito comparativo relativo alla coerenza percepita dagli utenti rispetto ai criteri di ricerca immessi evidenzia una leggera prevalenza del nuovo sistema (punteggio medio 4,29) rispetto al tradizionale (punteggio medio 3,79). Lo scarto maggiore si rileva nel gradimento da parte degli utenti 'non professionali' (utenti generalisti e melomani), mentre gli 'archivisti' mostrano una leggera preferenza per il sistema tradizionale (scarto medio di 0,50). Nel complesso i due sistemi risultano in grado di fornire risultati che sono ritenuti coerenti ai criteri impostati.

5.5.3.1.2 Q 1.2 – Efficacia delle ricerche

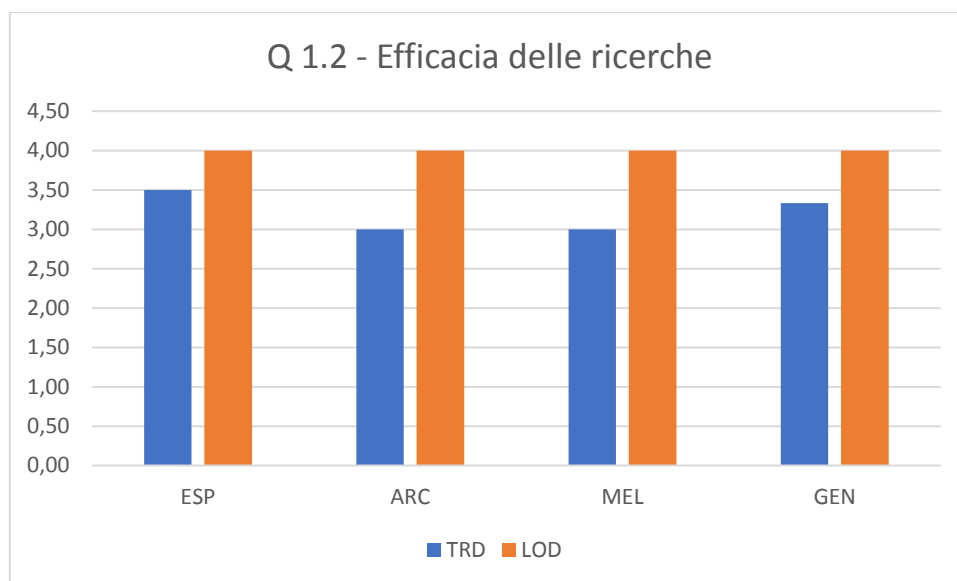


Diagramma 5.3: Q 1.2 - Efficacia delle ricerche

Il quesito comparativo relativo all'efficacia dei due sistemi nel reperire le informazioni d'interesse per gli utenti ha evidenziato una netta prevalenza del sistema basato sulle tecnologie semantiche. Corago LOD ha ottenuto un punteggio medio di 4,0 contro un 3,21 del sistema tradizionale. La preferenza per il nuovo approccio alla ricerca dei contenuti dell'archivio è espressa in modo concorde da tutte le tipologie di utente con scarti che vanno dallo 0,50 degli esperti di dominio, fino allo 1,0 di archivisti e melomani.

5.5.3.1.3 Q 1.3 – Performance

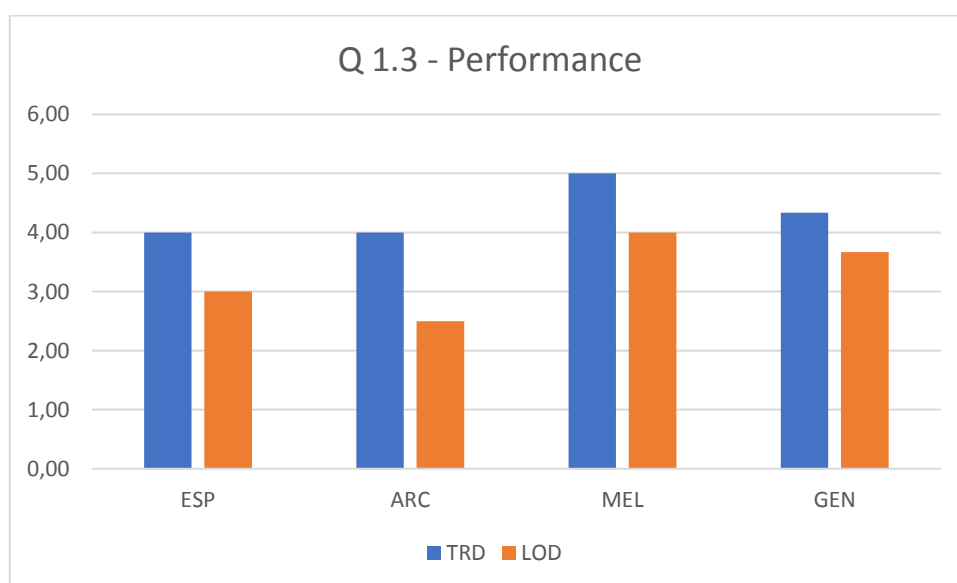


Diagramma 5.4: Performance

La valutazione delle performance complessive dei due sistemi evidenzia uno scarto a favore del sistema tradizionale con un punteggio medio di 4,33 contro un 3,29 ottenuto dal sistema sperimentale. Se quindi le performance percepite dagli utenti possono dirsi “buone” per il primo, per il secondo sono poco più che “sufficienti”. Evidentemente il sistema basato su tecnologie consolidate e su una architettura di sistema specificamente progettata per la sola consultazione¹⁸² costituisce ancora il benchmark rispetto a cui tendere nell’allestimento di applicazioni di consultazione di basi di conoscenza basate su grafi in formato RDF.

5.5.3.1.4 Q 1.4 – Raffinamenti delle ricerche

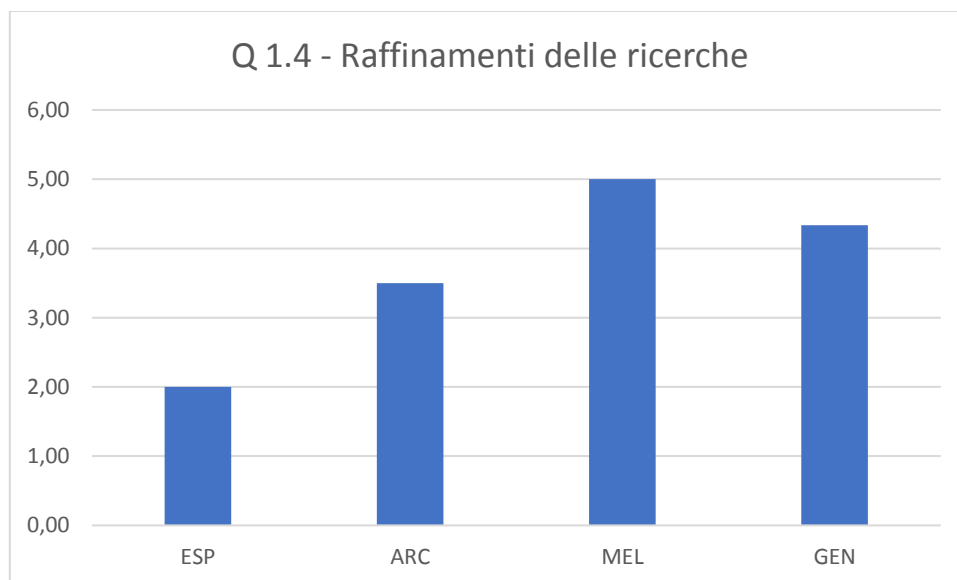


Diagramma 5.5: Raffinamenti delle ricerche

Il quesito voleva misurare la facilità con cui l’utente è stato in grado di ottenere risultati più aderenti alle sue aspettative utilizzando il sistema sperimentale. Complessivamente gli utenti hanno trovato relativamente “facile” raffinare le ricerche. Va però rilevata la significativa differenza delle valutazioni tra gli utenti professionali e gli altri. Se gli esperti di dominio reputano “difficile” raffinare le ricerche, gli altri ritengono più che “facile” riuscirci: il punteggio medio assegnato da GEN e MEL è rispettivamente 4,33 e 5,0. Se ne può derivare che il modello concettuale proposto all’utente dal sistema sperimentale risulta più intuitivo, o leggibile, e che la generalizzazione dei risultati risulta più efficace per l’utente con una minore conoscenza pregressa del dominio rispetto agli utenti più esperti.

¹⁸² A puro titolo informativo va riportato che il sistema di consultazione del *Corago* tradizionale non si basa sul modello relazionale impiegato dalla applicazione di catalogazione ed utilizzato come fonte per la produzione del *dataset* RDF, bensì su una sua versione denormalizzata in formato MySQL che alimenta esclusivamente l’applicazione web.

5.5.3.1.5 Q 1.5 – Utilità ricerca spazio-tempo

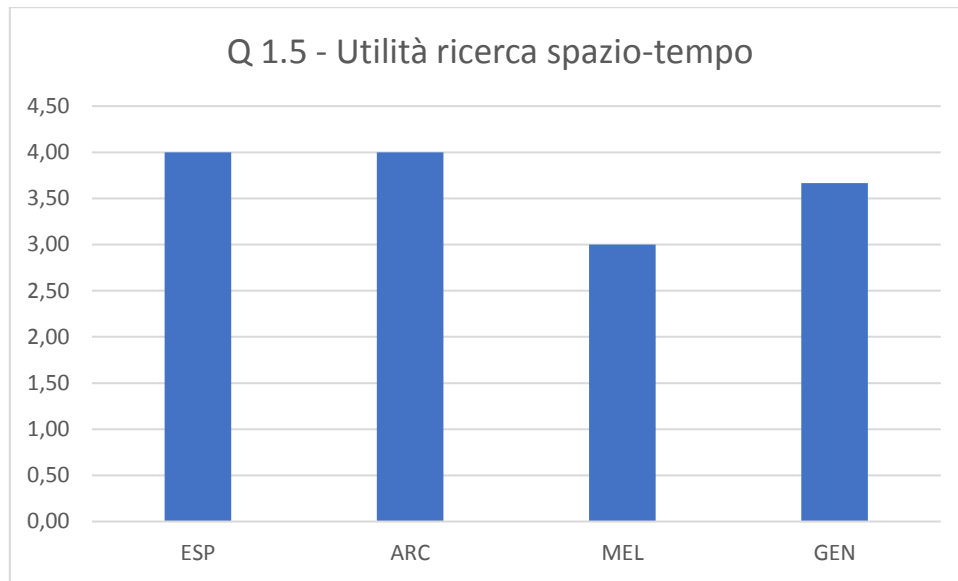


Diagramma 5.6: Utilità ricerca spazio-tempo

La valutazione dell'utilità delle funzionalità di ricerca spazio-tempo introdotte dal sistema sperimentale ha evidenziato un sostanziale gradimento da parte di tutte le tipologie di utente (punteggio medio 3,67) con una preferenza maggiore da parte degli utenti professionali (voto medio 4,0). Questo risultato sottolinea l'efficacia dell'impiego di spazio e tempo come dimensioni per accedere al patrimonio informativo dell'archivio, modalità di accesso resa possibile grazie alla normalizzazione della loro rappresentazione descritta nel paragrafo 3.1.

5.5.3.2 Efficacia nella rappresentazione dei contenuti

5.5.3.2.1 Q 2.1 – Leggibilità dei risultati

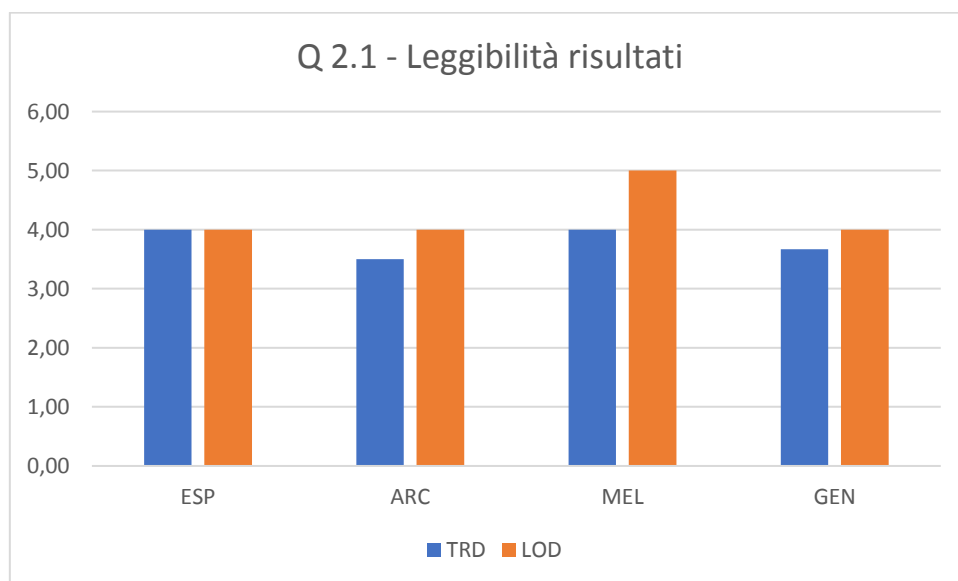


Diagramma 5.7: Q 2.1 - Leggibilità risultati

I risultati circa la 'leggibilità' degli elenchi di risultati proposti dai due sistemi sembrano indicare una leggera prevalenza del sistema sperimentale per il quale viene giudicata "molto" leggibile la presentazione proposta con un punteggio medio di 4.25 a fronte di un punteggio di 3.79 ottenuto dal sistema tradizionale. Lo scarto tra i due, rispetto alle diverse categorie d'utenza, va dalla sostanziale equivalenza percepita dagli utenti esperti del dominio ad una media di scarto pari allo 0,60 per le altre.

5.5.3.2.2 Q 2.2 – Utilità del contesto informativo

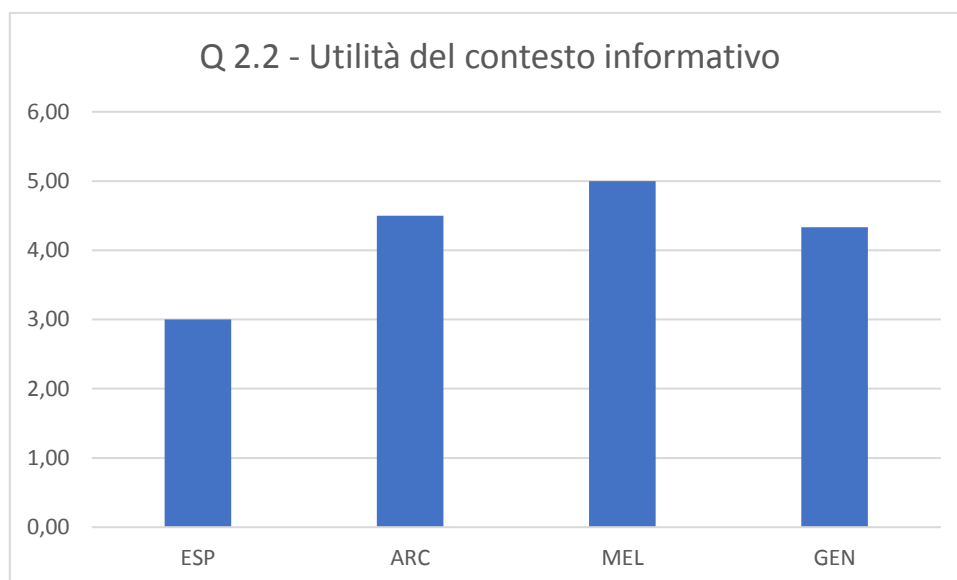


Diagramma 5.8: Q 2.2 - Utilità del 'contesto informativo'

La leggera prevalenza nel rappresentare in modo leggibile i risultati delle ricerche da parte del sistema sperimentale evidenziata dal quesito 2.1 (vedi diagramma 5.7), trova una potenziale motivazione nell'utilità percepita del cosiddetto 'contesto informativo' (vedi paragrafi 2.6 e 5.1) che viene giudicato 'molto' utile. In particolare, viene apprezzato dagli utenti non professionali con un punteggio medio di 4.67 (per altro anche gli archivisti/bibliotecari lo premiano con un 4.50). In base a questi risultati può dirsi confermata l'importanza del [REQ4] (vedi paragrafo 1.5.2) nel rispondere all'esigenza dell'utente di poter individuare facilmente i riferimenti tra il singolo risultato ed i contenuti in base ai quali è stato selezionato. In questo caso, i sistemi basati su modelli concettuali a semantica esplicita sembrano consentire di realizzare soluzioni applicative più efficaci.

5.5.3.2.3 Q 2.3 – Leggibilità delle schede

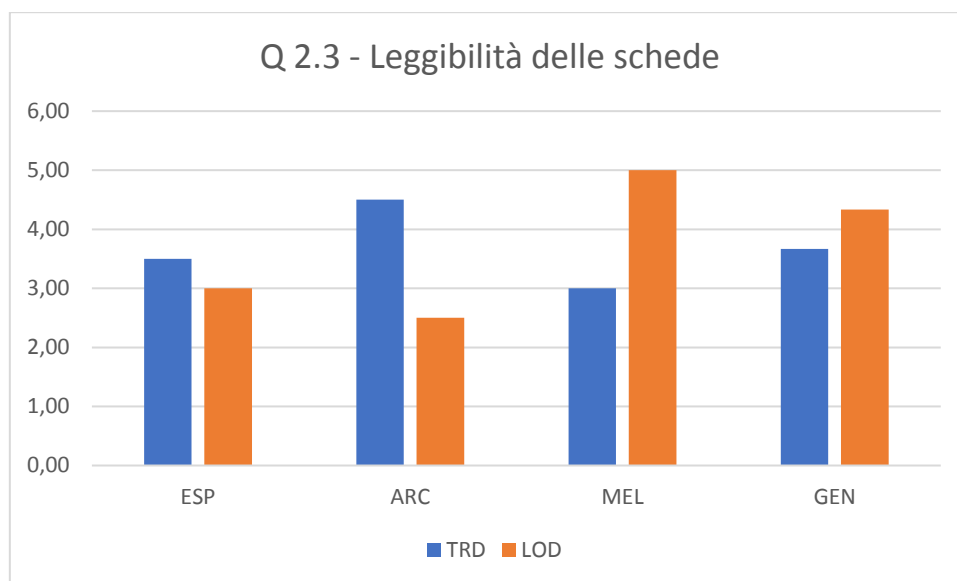


Diagramma 5.9: Q 2.3 - Leggibilità delle schede

La percezione della leggibilità delle schede delle singole entità proposte dai due sistemi ha invece mostrato una netta polarizzazione tra le due macro-categorie di utenza. Se il risultato medio è sostanzialmente comparabile tra i due sistemi: punteggio di 3.67 per il *Corago* tradizionale e 3.71 per il nuovo, gli utenti professionali hanno preferito il primo (con una media di 4.0) rispetto al secondo (media valutazioni di 2.75). Gli utenti non professionali hanno invece valutato il sistema sperimentale molto positivamente (media valutazioni di 4,67) ritenendo comunque 'abbastanza' leggibile il sistema tradizionale (con una media di 3.33). Il risultato sembra evidenziare la necessità di specializzare la composizione delle viste di presentazione dei contenuti in funzione delle aspettative dell'utenza. In questa prospettiva il modello di rappresentazione presentato in questo studio (vedi paragrafi 2.5 e 5.4) costituisce il punto di partenza per realizzare viste dedicate ad attori con aspettative funzionali diverse.

5.5.3.2.4 Q 2.4 – Utilità navigazione del grafo

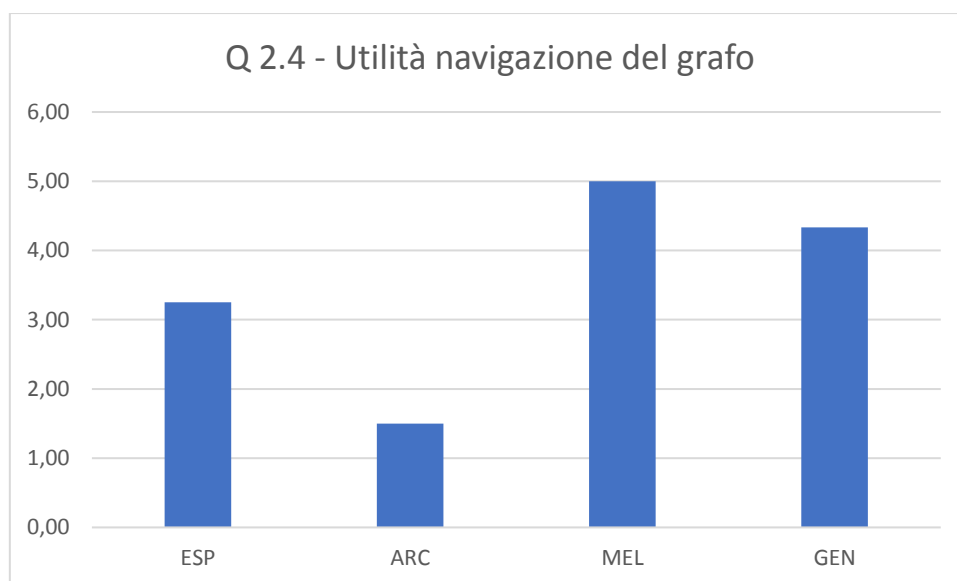


Diagramma 5.10: Q 2.4 - Utilità navigazione del grafo

Anche rispetto all'utilità della possibilità di navigare liberamente il grafo sembra evidenziarsi una diversa percezione tra le due macro-categorie di utenza. Se i non professionisti valutano 'molto' utile questo grado di libertà offerto dal sistema sperimentale (punteggio medio di 4.67), i professionisti lo ritengono 'poco' utile assegnando un punteggio medio di 2.38. Si può cercare di interpretare questo risultato come una diversa modalità di fruizione dei contenuti: mentre l'utente non professionista trova utile 'esplorare' i contenuti dell'archivio seguendo le relazioni tra le entità in modo contestuale alla fruizione; l'utenza professionale sembra prediligere l'accesso puntuale all'informazione cercata. Si può quindi ritenere che il modello di navigazione proposto dal sistema sperimentale consenta una fruizione dei contenuti meno vincolante che va incontro alle modalità di navigazione dell'utente non specialista. In questo senso sembrano raggiunti gli obiettivi di base individuati in fase di disegno della soluzione per la consultazione dell'archivio descritti nel paragrafo 1.5.1.

5.5.3.2.5 Q 2.5 – Efficacia della rappresentazione su cartografia

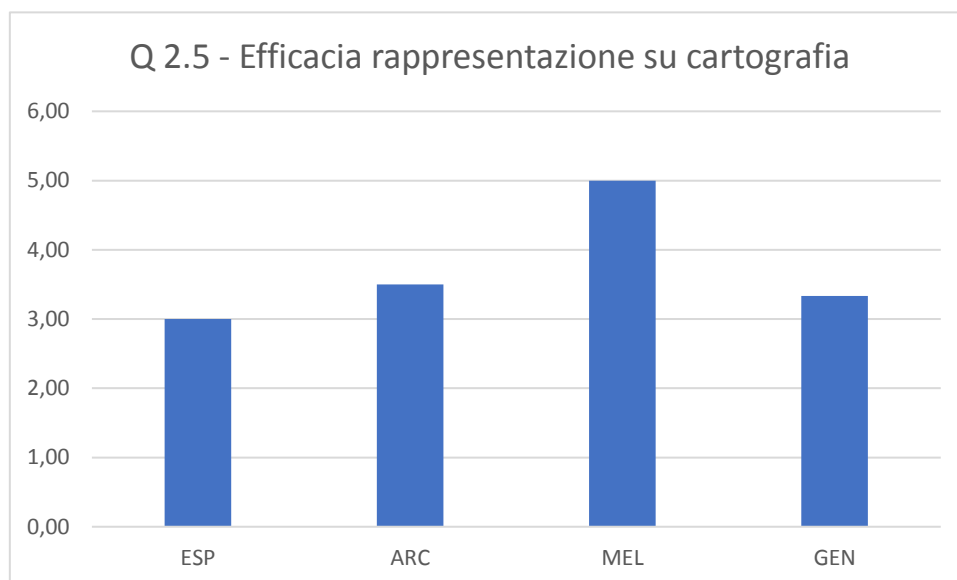


Diagramma 5.11: Q 2.5 - Efficacia rappresentazione su cartografia

La valutazione circa l'efficacia della rappresentazione delle informazioni attraverso la cartografia sembra invece riscuotere un sostanziale consenso ottenendo un punteggio medio di 3.71. Si tratta di una modalità di fruizione dei contenuti decisamente innovativa rispetto al dominio. Questo in parte può giustificare lo scarto tra la valutazione di 'abbastanza' utile da parte degli esperti di dominio e le valutazioni più positive degli altri tipi di utente che hanno assegnato una media di 3.94 corrispondente ad un giudizio di 'molto' utile. Si tratta di una modalità di fruizione dell'archivio da sviluppare e promuovere sia in termini funzionali che di disseminazione presso la comunità di riferimento.

5.5.4 Valutazione delle risposte aperte

La terza sezione del questionario, concepita per raccogliere un *feedback* aperto, è stata utilizzata dai soggetti per due finalità prevalenti: motivare ed estendere il giudizio espresso nelle due sezioni precedenti proponendo quelle migliorie che avrebbero portato ad una valutazione migliore del nuovo sistema; proporre migliorie e nuove funzionalità non trattate dal questionario. Si riporta una sintesi di quelle ritenute più significative di entrambe le tipologie.

5.5.4.1.1 Integrazione ambiti rilevati dal questionario

Gli utenti N1_ARC e N6_ESP vengono sottolineati gli aspetti relativi alle performance ed all'usabilità del nuovo sistema. Entrambi rilevano la necessità di raffinare ulteriormente il *layout* delle schede di presentazione delle singole entità.

Gli utenti N1_ARC e N3_ARC rilevano la necessità di migliorare gli strumenti contestuali per il raffinamento delle ricerche. In particolare, N1_ARC suggerisce l'introduzione delle faccette (*faceted browsing*) come strumento per il raffinamento dei risultati delle ricerche.

Gli utenti N2_MEL e N4_GEN rilevano la necessità di rendere più evidenti le logiche di navigazione tra le diverse pagine.

5.5.4.1.2 Altri ambiti non coperti dal questionario

Gli utenti N2_MEL, N3_ARC e N4_GEN rilevano l'opportunità di migliorare l'organizzazione dei contenuti e la veste grafica della home page.

5.5.5 Considerazioni complessive sui risultati ottenuti

Nel complesso i risultati dell'indagine condotta attraverso il questionario, per quanto circoscritta ed improntata ad una valutazione prevalentemente qualitativa, evidenziano una sostanziale efficacia delle soluzioni individuate per il soddisfacimento dei requisiti del sistema sperimentale di consultazione.

Una volta verificata la fattibilità della loro trasposizione in un modello concettuale a semantica esplicita si è posto il problema di verificare la possibilità di realizzare una applicazione di consultazione che rispondesse, in termini quantomeno equivalenti alla soluzione tradizionale, alle aspettative di accesso al patrimonio informativo dell'archivio nella sua nuova forma.

I risultati dei quesiti Q 1.1, Q 1.2 e Q 1.4 sembrano confermare che, in termini di efficacia nella ricerca dei contenuti, il nuovo sistema sia in grado di eguagliare l'approccio tradizionale. Vengono introdotti al contempo elementi di maggiore flessibilità (quesito Q 2.4) ed esaustività (quesito Q 2.2) nell'accesso alle informazioni.

Se dal punto di vista implementativo restano da risolvere aspetti legati alle performance (quesito Q 1.3) e di rappresentazione dei contenuti (quesito Q 2.3), le nuove funzionalità di ricerca spazio-tempo e di visualizzazione su cartografia (entrambe introdotte grazie all'inclusione di contenuti provenienti da fonti LOD) hanno ottenuto un riscontro decisamente positivo (quesiti Q 1.5 e Q 2.5).

Le indicazioni ricavate dal quesito a risposta aperta (sezione 3 del questionario) sembrano indicare l'interesse per un ulteriore affinamento del sistema sperimentale in grado di portarlo ad un livello di maturità tale da poter sostituire la soluzione tradizionale. In questa prospettiva i risultati ottenuti dal sistema sperimentale sul piano funzionale possono costituire, alla luce del confronto con gli utenti, una valida base di partenza.

6 Conclusioni

6.1 Rappresentazione dei contenuti

Nel capitolo 2 è stato illustrato il percorso che ha consentito di verificare come attraverso l'impiego delle ontologie CIDOC CRM e FRBR Object Oriented sia possibile rappresentare efficacemente la conoscenza relativa alla storia del melodramma raccolta nell'archivio *Corago*. Per quanto efficace, l'impiego delle due ontologie ha comportato l'introduzione di una significativa complessità della morfologia del grafo RDF ottenuto a valle della trasposizione dei contenuti. Ne è derivata la necessità di elaborare delle diverse modalità di rappresentazione dell'informazione che risultassero efficaci rispetto alle esigenze dell'utente finale.

A questo scopo sono stati introdotti due strumenti specifici: un modello per l'indicizzazione dei contenuti testuali ed un modello di astrazione concettuale orientato alla navigazione dei contenuti della base di conoscenza. Il primo ha permesso di realizzare una funzione di ricerca libera testuale in grado di collocare i risultati individuati all'interno del 'contesto informativo' corrispondente ai criteri di ricerca. Il sistema sperimentale realizzato è così in grado di produrre l'insieme delle relazioni che intercorrono tra i contenuti testuali relativi al risultato e i termini impiegati. In questo modo, ciascun risultato è direttamente messo in relazione ai parametri utilizzati così da consentire una valutazione sia della sua coerenza sia della rilevanza rispetto agli obiettivi che l'utente si era prefissato (vedi paragrafo 2.6).

L'introduzione di un modello concettuale orientato alla navigazione ha permesso di attestare la dereferenziazione delle risorse interamente sul modello formalizzato nell'ontologia di dominio *Corago SM*. L'esperto di dominio, grazie all'introduzione di questo *layer*, è in grado di definire la rilevanza dei contenuti in modo da produrre una loro rappresentazione coerente ed efficace. Questo ha consentito di superare la complessa articolazione del grafo RDF producendo delle viste utente costruite in una prospettiva squisitamente funzionale. L'aver formalizzato questa competenza relativa alle modalità di rappresentazione delle entità del dominio attraverso il linguaggio OWL fa sì che le competenze di dominio siano riutilizzabili dalla comunità di riferimento (vedi paragrafo 2.5). In questo senso l'interoperabilità del *dataset* risulta ampliata.

L'adozione di ontologie non specialistiche come il CIDOC CRM per rappresentare l'informazione relativa al melodramma permette di superare i limiti di interoperabilità tipici dei sistemi basati su modelli concettuali proprietari. Nella sperimentazione si è però evidenziato come sia comunque necessario individuare un modello concettuale articolato su più livelli capaci di rappresentare in modo più efficace il dato dal punto di vista dell'utente. Ciascun livello potrà così rispondere a requisiti di natura diversa. Lo strato del livello di base sarà orientato all'interoperabilità dell'informazione tra sistemi rispondendo così ai principi Linked Data. Al di sopra del livello base si dovranno elaborare strati intermedi in grado di rispondere alle specifiche esigenze funzionali del singolo dominio.

6.2 Analisi dei contenuti

La migrazione dei contenuti verso un modello concettuale a semantica esplicita e la conseguente organizzazione dei dati in forma di grafo hanno consentito lo sviluppo di nuove modalità di accesso. Sono state individuate dimensioni di analisi in grado di far affiorare nessi tra i contenuti altrimenti difficilmente individuabili nella precedente organizzazione delle informazioni. La sperimentazione relativa all'impiego del personaggio come dimensione di analisi ha dimostrato l'efficacia nell'individuazione di relazioni tra i testi letterari partendo da informazioni già presenti nella base di conoscenza.

Le due metodologie di estrazione dell'informazione sperimentate (l'analisi delle tradizioni dei testi e l'estrazione delle relazioni interpersonali tra personaggi) evidenziano l'efficacia nell'applicazione di metodologie sia di tipo supervisionato che non, a partire da un modello concettuale comune. Nel caso dei personaggi è stato possibile individuare su un sottoinsieme del repertorio oltre duecento casi di relazioni tra testi (vedi paragrafo 3.2.3); nel secondo caso si è riusciti a classificare oltre mille occorrenze di relazioni tra i personaggi (vedi paragrafo 4.2.4.1). Sia i modelli descrittivi utilizzati per l'analisi che i risultati in termini di contenuti estratti sono stati fatti confluire nel patrimonio informativo dell'archivio.

In via preliminare, si può affermare che la formalizzazione della conoscenza e l'applicazione di metodologie per l'estrazione dell'informazione siano strumenti utilmente impiegabili per l'avanzamento della conoscenza di dominio. D'altra parte, è opportuno sottolineare che le analisi effettuate non avrebbero potuto essere svolte in assenza di un modello concettuale flessibile e di strumenti per la trattazione automatica del linguaggio naturale. Questo non solo in considerazione della quantità di informazione processata, ma anche per i metodi di indagine che l'adozione di questi strumenti ha sollecitato. La necessità di produrre una definizione formale delle entità coinvolte nell'analisi e la conseguente organizzazione dell'informazione porta ad un processo ciclico che, ad ogni iterazione, vede arricchirsi non solo la conoscenza estratta ma anche affinarsi il modello concettuale destinato a raccoglierla.

Assieme ai contenuti si arricchisce la capacità di gestione e rappresentazione dell'informazione in un percorso in cui l'esperto di dominio è contemporaneamente soggetto agente dell'indagine e fruitore dei risultati. In questo senso la sperimentazione effettuata costituisce uno spunto per approfondire l'impiego di queste tecnologie come strumenti in grado di produrre un avanzamento delle metodologie di trattamento dell'informazione storica e documentale anche in altri ambiti disciplinari.

6.3 Accesso ai contenuti

L'impianto del modello FRBR alla base dell'archivio *Corago* è improntato alla fruizione del dato da parte degli utenti finali del sistema. L'orientamento funzionale determina sia la tipologia di dati raccolti che la loro organizzazione. Questo a prescindere dagli strumenti col quale il sistema di catalogazione è stato realizzato. Si è constatato anche in questo studio come l'adozione di una specifica tecnologia non risulti completamente trasparente. L'introduzione della sovrastruttura semantica porta ad una articolazione dell'informazione che svincola il dato dal modello fisico. Questo offre la possibilità di produrre rappresentazioni diverse degli stessi contenuti. Si è potuto verificare come sia possibile realizzare un sistema di consultazione che affianchi, alle tradizionali funzioni di ricerca e navigazione, nuove modalità di fruizione del patrimonio informativo dell'archivio. Siano esse funzionalità di ricerca trasversali, navigazioni reticolari o nuove metafore visuali. L'obiettivo resta quello di aumentare la fruibilità dei contenuti e, nella più ampia prospettiva di una integrazione con l'ecosistema dei Linked Data, di espandere la navigazione verso altre basi di conoscenza. Il riscontro rilevato attraverso il sondaggio di gradimento del sistema sperimentale ha evidenziato l'efficacia delle nuove funzionalità introdotte, ma anche la necessità di raggiungere prestazioni equivalenti ai sistemi tradizionali (vedi paragrafo 5.5.3).

Nel complesso, lo studio evidenzia come l'insieme degli strumenti sperimentati offrano significative potenzialità. Tali strumenti sono un ineludibile termine di paragone per i futuri sistemi per la gestione dell'informazione storica e documentale soprattutto in un contesto in cui il dato nasce sempre più frequentemente in forma digitale e distribuita (Calhoun, 2011, p. 38). L'evoluzione verso sistemi in grado di scambiare informazioni sulla base di un allineamento tra la semantica dei rispettivi modelli si pone inoltre come possibile soluzione alla sostenibilità della gestione di contenuti multiformi e complessi come, nel nostro caso, quelli relativi alla storia del melodramma. Il successo di questi strumenti presso la comunità degli

studiosi e del pubblico dipenderà in larga parte dalla capacità delle istituzioni preposte alla conservazione e valorizzazione del patrimonio culturale di promuoverne l'impiego d'intesa con la comunità scientifica e professionale di riferimento.

Ringraziamenti

Voglio ringraziare il mio supervisore, il professor Angelo Pompilio, per avermi dato la possibilità di svolgere questo studio, per l'infinita pazienza nel condividere le peregrinazioni di una riflessione ormai decennale e per avermi guidato in questo lavoro che ne rappresenta il punto di approdo.

Vorrei ringraziare il collegio ed i colleghi del Corso di Dottorato per gli stimoli, i consigli e la condivisione del proprio entusiasmo.

Devo ringraziare la dott.ssa Marta Zaetta per i preziosi consigli nell'allestimento del questionario di valutazione.

Ed in ultimo mia moglie ed i miei ragazzi che semplicemente hanno reso possibile tutto questo.

Prefissi utilizzati

Nel corso della trattazione tutti i riferimenti alle entità delle diverse ontologie citate sono riportati per brevità nella forma contratta. Di seguito vengono riportati i *namespace* utilizzati ed i relativi URI.

corago="http://corago.unibo.it/sm/"
owl="http://www.w3.org/2002/07/owl#"
rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xml="http://www.w3.org/XML/1998/namespace"
xsd="http://www.w3.org/2001/XMLSchema#"
crm="http://erlangen-crm.org/current/"
rdfs="http://www.w3.org/2000/01/rdf-schema#"
skos="http://www.w3.org/2004/02/skos/core#"
udep="http://universaldependencies.org/u/dep#"
frbroo="http://erlangen-crm.org/efrbroo/"
famtree="http://www.co-ode.org/roberts/family-tree.owl#"
geovocab="http://geovocab.org/geometry#"
geosparql="http://www.opengis.net/ont/geosparql#"
relationship="http://purl.org/vocab/relationship/"
ud="http://universaldependencies.org/u/dep#"
yago="http://dbpedia.org/class/yago/"
dbpedia="http://dbpedia.org/ontology/"
geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
dcterms="http://purl.org/dc/terms/"

Allegati digitali

Costituiscono parte integrante della tesi le seguenti risorse digitali pubblicate attraverso le rispettive piattaforme di conservazione:

1. il Dataset Corago LOD: Bonora Paolo, & Pompilio Angelo. (2019). Corago LOD raw dataset (Version 20190824) [Data set], Zenodo, DOI: <http://doi.org/10.5281/zenodo.3377586>;
2. l'ontologia Corago SM: progetto sulla piattaforma GitHub reperibile all'indirizzo: <https://github.com/paolobonora/Corago-SM>;
3. le relazioni tra le opere: Bonora, Paolo (2019): Relazioni opere da analisi personaggi.xlsx, figshare Dataset, FigShare, DOI: <https://doi.org/10.6084/m9.figshare.9756788>;
4. le annotazioni alle didascalie dei personaggi: Bonora, Paolo (2019): NLP annotations of characters descriptions, figshare, Dataset, FigShare, DOI: <https://doi.org/10.6084/m9.figshare.7880489>;
5. i questionari di valutazione: Bonora, Paolo (2019): Corago LOD's users satisfaction survey, figshare, Dataset, DOI: <https://doi.org/10.6084/m9.figshare.9466505>.

Appendici

A. Mappe concettuali per la trasposizione dei contenuti

L'appendice illustra in modo sintetico le corrispondenze (*mapping*) tra le principali entità del modello relazionale Corago e le classi di destinazione del modello concettuale costituito dalle due ontologie di riferimento (CIDOC CRM e FRBRoo) e dall'ontologia di dominio (Corago SM).

Le corrispondenze utilizzate tra i due modelli vengono illustrate attraverso le mappe concettuali suddivise per area tematica. Le mappe riportano le classi coinvolte e le relative relazioni di ciascuna delle dieci aree. Nelle tabelle riassuntive che accompagnano ciascuna mappa vengono riportate le associazioni tra campi di origine e proprietà di destinazione.

Per la formulazione completa della articolazione del *mapping* si rinvia al relativo file in linguaggio D2RQ-ML pubblicato attraverso il progetto GitHub “Corago-Mapping” all'indirizzo: <https://github.com/paolobonora/Corago-Mapping>.

A.1. Mapping dei Responsabili

L'attribuzione di responsabilità è una delle aree a maggiore densità di informazione del modello di origine. È caratterizzata dalla forte tipizzazione del ruolo di responsabilità rispetto all'entità associata: relazione spesso arricchita da attributi che ne specificano natura e rilevanza.

Il diagramma rappresenta la parte relativa alle anagrafiche rinviando le attribuzioni di responsabilità ai singoli diagrammi della entità oggetto.

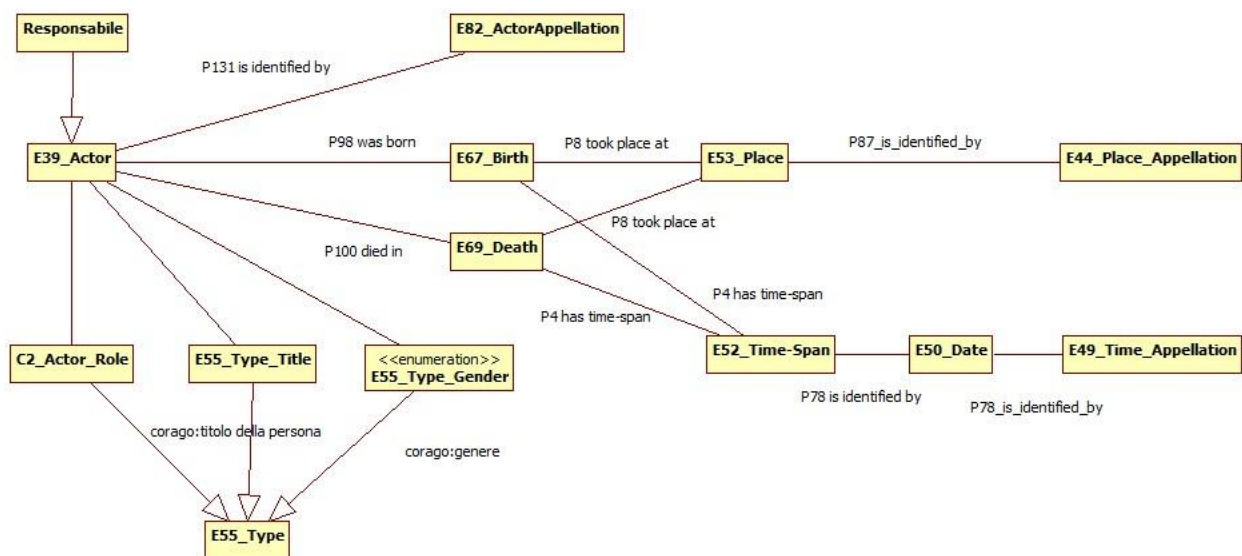


Diagramma A.1: Mapping dell'entità Responsabile

Nella definizione del *mapping* si è cercato di seguire quanto più possibile le linee guida indicate in (Doerr, 2009), in particolare quelle relative al *mapping* per entità derivate dai modelli FRBR-ER e FRAD.

Si è presentata qualche difficoltà per specializzazione degli URI per le tre sottoclassi E55_Type_Role, E55_Type_Title, E55_Type_Gender essendo necessario attestarle tutte su E55_Type, ma partendo da attributi distinti.

Il modello relazionale utilizza come strumento di tipizzazione la tabella TADE (acronimo di ‘tabella delle definizioni’) alla quale quasi tutte le entità vi ricorrono per specificare caratteristiche proprie o di relazione. Qui, come altrove, si è cercato di attestare queste tipizzazioni su E55_Type o sue specializzazioni in Corago SM. I diagrammi, dove rilevante, riportano come attributo della classe derivata l’indicazione del codice identificativo del Tipo TADE che qualifica la classe generata (il dominio delle sue istanze è definito dall’insieme dei record della tabella TADE che hanno quel Tipo [riportato nel campo “TADETIP”]).

Una peculiarità è rappresentata dal ruolo di responsabilità generica (qualifica autoriale assoluta in Corago). È una responsabilità non derivata da una relazione specifica con un’opera: per questa si è scelto di generare tante istanze di E55_Type normalizzando il campo RESPONS.COD_TIPO_R.

Nella formazione degli URI delle classi intermedie E67 Birth e E6 Death, come per la classe ausiliaria E52 Time-Span necessarie per esprimere la semantica della dimensione temporale, si è scelto di introdurre una desinenza specifica all’URI dell’istanza della classe foglia del ramo.

A.1.1. Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
Responsabile	crm:E39_Actor	RESPONS/@@RESPONS.COD_RESP urlify@@	rdfs:label	RESPONS.COGNOME_R RESPONS.NOME_R

A.1.2. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	PersonTypeGender	RESPONS/GENDER&@@RESPONS.COD_RESP urlify@@	crm:E55_Type	crm:P2_has_type
Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	PersonTypeRole	TADE/RES&@@TADE.ID_TADE urlify@@	corago:C2_Actor_Role	crm:P2_has_type
Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	PersonTypeTitle	RESPONS/TITLE&@@RESPONS.COD_RESP urlify@@	crm:E55_Type	crm:P2_has_type
Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	ActorAppellation	RESPONS/APPL/@@RESPONS.COD_RESP urlify@@	crm:E82_Actor_Appellation	crm:P131_is_identified_by
Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	PersonBirthDate	RESPONS/NAS/@@RESPONS.COD_RESP urlify@@	crm:E67_Birth	crm:P98i_was_born
Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	PersonDeathDate	RESPONS/MOR/@@RESPONS.COD_RESP urlify@@	crm:E69_Death	crm:P100i_died_in
ActorAppellation	RESPONS/APPL/@@RESPONS.COD_RESP urlify@@	crm:E82_Actor_Appellation	Responsabile_AppellationAltreFormeTipo	TADE/RES/PARENT_FORM	crm:E55_Type	crm:P2_has_type

A.2. Mapping dei Brani

Il diagramma A.2 rappresenta l'insieme delle classi generate a partire dall'entità Brano. Il diagramma include oltre al livello di Work ed Expression anche le relative ramificazioni (ad es. verso la classe frbroo:F31_Performance).

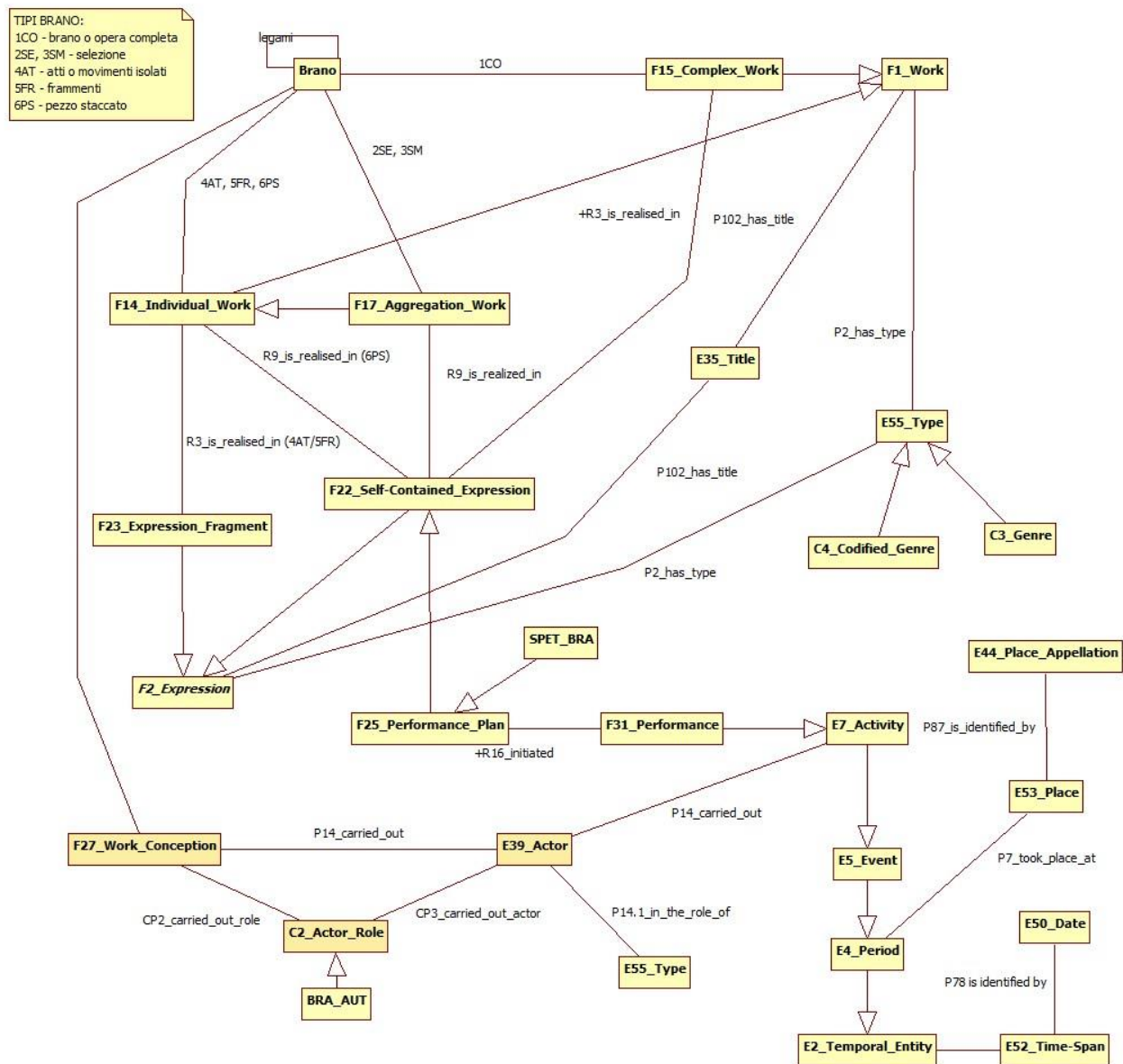


Diagramma A.2: Mapping dell'entità Brano

Nel modello originario dell'archivio *Corago* i due piani FRBR relativi al Work e alla sua Expression non sono fisicamente distinti. Di conseguenza si è dovuto procedere alla derivazione delle istanze di frbroo:F1_Work e frbroo:F2_Expression a partire dalla stessa tabella BRANI. Nella generazione delle istanze si è proceduto utilizzando la codifica del tipo di brano di *Corago* per creare istanze delle sottoclassi specializzate di frbroo:F1_Work in base alla configurazione riportata nella tabella A.2.

Valore TIPO_BRANO	Classificazione TADE	Classificazione FRBRoo
1CO	brano o opera completa	F15 Complex Work
2SE	selezione	F17 Aggregation Work
3SM	selezione miscellanea	F17 Aggregation Work
4AT	atti o movimenti isolati	F14 Individual Work
5FR	frammenti	F14 Individual Work
6PS	pezzo staccato	F14 Individual Work

Tabella A.2: Associazioni tra tipi brani Corago e classi FRBRoo

L'attestazione delle tipologie di brano è stata fatta adottando un meccanismo di *mapping* che consente di gestire il polimorfismo della classe Brano nelle diverse specializzazioni previste in FRBRoo. Utilizzando il campo BRANI.TIPO_BRANO si è creata una tabella di transcodifica della proprietà *rdf:type* della classe nelle classi FRBRoo di destinazione.

Per ciascuna istanza di specializzazione del tipo `frbroo:F1_Work`, viene creata una ulteriore istanza di specializzazione di `frbroo:F2_Expression`. Per tutti quei Brani per cui è presente un record in SPET_BRA vengono create tante istanze di `frbroo:F25_Performance_Plan` per ogni relazione Brano-Spettacolo (con URI ID_SPET). Per i BRANI privi di record in SPET_BRA viene creata una istanza di sottoclassi di `frbroo:F2_Expression` secondo la seguente tabella A.3.

Valore TIPO_BRANO	Tipo di Work	Tipo di Expression
1CO	F15 Complex Work	F22 Self-Contained Expression
2SE	F17 Aggregation Work	F22 Self-Contained Expression
3SM	F17 Aggregation Work	F22 Self-Contained Expression
4AT	F14 Individual Work	F23 Expression Fragment
5FR	F14 Individual Work	F23 Expression Fragment
6PS	F14 Individual Work	F22 Self-Contained Expression

Tabella A.3: Associazioni tra tipi brani Corago e classi specializzate di F1_Work e F2_Expression

Nella gestione dello *switch* condizionale in funzione del valore dell'attributo è stato impiegato il costrutto offerto dal linguaggio D2RQ-ML: *d2rq:TranslationTable* che consente di definire una proprietà del nodo creato a partire dal valore di un campo del record di origine dell'entità. Nel nostro caso è stato utilizzato per valorizzare la proprietà *rdf:type* del `BranoType` poi associato all'entità Brano. Si riporta il codice utilizzato a scopo illustrativo nel listato seguente A.1.

```

# Gestione Brani (BRANI)
# Classe Brano a livello F1_Work
map:Brano a d2rq:ClassMap;
    d2rq:dataStorage map:database;
    d2rq:uriPattern "BRANI/@@BRANI.COD_BRANO@@";
.

# Gestione del polimorfismo per la classe Brano a livello F1_Work
map:BranoType a d2rq:PropertyBridge;
    d2rq:property rdf:type;
    d2rq:uriColumn "BRANI.TIPO_BRANO";
    d2rq:translateWith map:BranoTypeTable;
    d2rq:belongsToClassMap map:Brano
.

# Configurazione switch polimorfismo a livello F1_Work
map:BranoTypeTable a d2rq:TranslationTable;
    d2rq:translation [ d2rq:databaseValue "1CO"; d2rq:rdfValue frbroo:F15_Complex_Work; ];
    d2rq:translation [ d2rq:databaseValue "2SE"; d2rq:rdfValue frbroo:F17_Aggregation_Work; ];
    d2rq:translation [ d2rq:databaseValue "3SM"; d2rq:rdfValue frbroo:F17_Aggregation_Work; ];
    d2rq:translation [ d2rq:databaseValue "4AT"; d2rq:rdfValue frbroo:F14_Individual_Work; ];
    d2rq:translation [ d2rq:databaseValue "5FR"; d2rq:rdfValue frbroo:F14_Individual_Work; ];
    d2rq:translation [ d2rq:databaseValue "6PS"; d2rq:rdfValue frbroo:F14_Individual_Work; ];
.

```

Listato A.9: Esempio di codice in linguaggio D2RQ-ML

L'altro ambito rilevante di questa area è la relazione tra Opere e Autori (frbroo:F1_Work e crm:E39_Autor). Viene definita in *Corago* dalla tabella BRA_AUT con filtro PRINCIP=1. Nella trasposizione la relazione deve essere mediata da una istanza di frbroo:F27_Work_Conception associata con una frbroo:R16_was_initiated_by al frbroo:F1_Work e con una proprietà corago:CP2_carried_out_role ad una istanza corago:C2_Actor_Role a sua volta associata attraverso una corago:CP3_carried_out_actor alla relativa istanza di crm:E39_Actor. In questo caso si è reso necessario l'introduzione della classe corago:C2_Actor_Role proprio per specificare il tipo di responsabilità.

A.2.1. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
AmbientazioneAnnoBran	BRANI/AMBIENTAZIONE/ANNO/@@BRANI.COD_BRANO urlify@	crm:E52_Time-Span	AmbientazioneAnn	BRANI/AMBIENTAZIONE/ANNOA/MBE/@@BRANI.COD_BRANO urlify@@	crm:E49_TimeAppellation	crm:P78_is_identified_by
AmbientazioneAnnoBran	BRANI/AMBIENTAZIONE/ANNO/@@BRANI.COD_BRANO urlify@	crm:E52_Time-Span	AmbientazioneAnnFine	BRANI/AMBIENTAZIONE/ANNO/2/@@BRANI.COD_BRANO urlify@	crm:E62_String	crm:P80_end_is_qualified_by

AmbientazioneAnnoBran	BRANI/AMBIENTAZIONE/ANNO/@@BRANI.COD_BRANO urlify@@	crm:E52_Time-Span	AmbientazioneAnnolnizio	BRANI/AMBIENTAZIONE/ANNO/1/@@BRANI.COD_BRANO urlify@@	crm:E62_Strin	crm:P79_beginning_is_qualified_by
ConcezioneBran	BRANI/CON&@@BRANI.COD_BRANO urlify@@	frbroo:F27_Work_Conception	DataConcezioneBran	BRANI/ANNO_COMP/@@BRANI.COD_BRANO urlify@@	crm:E52_Time-Span	corago:CP14_conception_date
ConcezioneBran	BRANI/CON&@@BRANI.COD_BRANO urlify@@	frbroo:F27_Work_Conception	DataEsecuzioneBran	BRANI/ANNO_1ESEC/@@BRANI.COD_BRANO urlify@@	crm:E52_Time-Span	corago:CP13_first_performance_date
ConcezioneBran	BRANI/CON&@@BRANI.COD_BRANO urlify@@	frbroo:F27_Work_Conception	EdificioConcezioneBran	BRANI/EDIFICIO/@@BRANI.EDIF_1ESEC urlify@@	crm:E19_Physical_object	corago:CP12_first_performance_location
ConcezioneBran	BRANI/CON&@@BRANI.COD_BRANO urlify@@	frbroo:F27_Work_Conception	LuogoConcezioneBran	BRANI/LUOGO/@@BRANI.LOC_1ESEC urlify@@	frbroo:F9_Place	corago:CP11_first_performance_place
ConcezioneBran	BRANI/CON&@@BRANI.COD_BRANO urlify@@	frbroo:F27_Work_Conception	RuoloResponsBran	BRA_AUT/@@BRA_AUT.ID_BRA_AUT urlify@@	corago:C18_Author_Role	corago:CP2_carried_out_role
DataConcezioneBran	BRANI/ANNO_COMP/@@BRANI.COD_BRANO urlify@@	crm:E52_Time-Span	DataConcezioneBranAPP	DATAAP/T_A_COMP/@@BRANI.T_A_COMP urlify@@	corago:C16_Approximate_Time_Appellation	crm:P78_is_identified_by
DataConcezioneBran	BRANI/ANNO_COMP/@@BRANI.COD_BRANO urlify@@	crm:E52_Time-Span	DataConcezioneBranTA	DATA/@@BRANI.ANNO_COMP urlify@@	crm:E49_Time_Appellation	crm:P78_is_identified_by
DataEsecuzioneBran	BRANI/ANNO_1ESEC/@@BRANI.COD_BRANO urlify@@	crm:E52_Time-Span	DataEsecuzioneBranAPP	DATAAP/T_A_1ESEC/@@BRANI.T_A_1ESEC urlify@@	corago:C16_Approximate_Time_Appellation	crm:P78_is_identified_by
DataEsecuzioneBran	BRANI/ANNO_1ESEC/@@BRANI.COD_BRANO urlify@@	crm:E52_Time-Span	DataEsecuzioneBranTA	DATA/@@BRANI.ANNO_1ESEC urlify@@	crm:E49_Time_Appellation	crm:P78_is_identified_by
Legame	LEGAMI/@@TADE.TADETIP@@/@@LEGAMI.COD_BRANO@@-@@LEGAMI.COD_BRANOL urlify@@	corago:C5_Link	NoteLegame	LEGAMI/NOTE/@@LEGAMI.NOT_E_LEG urlify@@	crm:E62_Strin	crm:P3_has_note
LuogoConcezioneBran	BRANI/LUOGO/@@BRANI.LOC_1ESEC urlify@@	frbroo:F9_Place	LuogoConcezioneBranPA	LUOGO/@@BRANI.LOC_1ESEC urlify@@	crm:E44_Place_Appellation	crm:P1_is_identified_by
PianoBran	SPET_BRA/@@SPET_BRA.ID_SPE_T_BRA urlify@@	frbroo:F25_Performance_Plan	Genere	TADE/GEN&@@TADE.ID_TADE urlify@@	corago:C3_Genre	crm:P2_has_type

PianoBr ano	SPET_BRA/@@SPET_BRA.ID_SPE T_BRA urlify@@	frbroo:F2 5_Perfor mance_Pl an	PriorityPi anoBrano	SPET_BRA/NPRO/@@SPET_BRA.I D_SPE_T_BRA urlify@@/@@SPET _BRA.NPRO@@	crm:E60_Num ber	corago:CPAN 1_priority
PianoBr ano	SPET_BRA/@@SPET_BRA.ID_SPE T_BRA urlify@@	frbroo:F2 5_Perfor mance_Pl an	PianoBra noNote	SPET_BRA/NOTE/@@SPET_BRA.I D_SPE_T_BRA urlify@@	crm:E62_Strin g	crm:P3_has_ note
PianoBr ano	SPET_BRA/@@SPET_BRA.ID_SPE T_BRA urlify@@	frbroo:F2 5_Perfor mance_Pl an	PianoBra noTipo	SPET_BRA/FORMA/@@SPET_BR A.ID_SPE_T_BRA urlify@@	crm:E55_Type	crm:P2_has_ type
PianoBr ano	SPET_BRA/@@SPET_BRA.ID_SPE T_BRA urlify@@	frbroo:F2 5_Perfor mance_Pl an	Productio nRelease	TADE/TAL&@@TADE.ID_TADE u rlify@@	corago:C9_Pr oduction_Rele ase	crm:P2_has_ type
PianoBr ano	SPET_BRA/@@SPET_BRA.ID_SPE T_BRA urlify@@	frbroo:F2 5_Perfor mance_Pl an	TitoloPian oBrano	TITOLI/SPET_BRA_TIT/@@SPET_ BRA_TIT.ID_SPE_T_BRA urlify@@ /@@SPET_BRA_TIT.NPRO urlify @@	crm:E35_Title	crm:P102_ha s_title
RuoloRe spBrano	BRA_AUT/@@BRA_AUT.ID_BRA_ AUT urlify@@	corago:C 18_Autho r_Role	Espressio neDocum ento	SWT_DOC_BRA/@@SWT_DOC_B RA.ID_DOC_BRA urlify@@	frbroo:F24_Pu blication_Expr ession	corago:CP10 _documente d_in
RuoloRe spBrano	BRA_AUT/@@BRA_AUT.ID_BRA_ AUT urlify@@	corago:C 18_Autho r_Role	PriorityRe sponsabil e	BRA_AUT/@@BRA_AUT.ID_BRA_ AUT urlify@@/@@BRA_AUT.NP RO urlify@@	crm:E60_Num ber	corago:CPAN 1_priority
RuoloRe spBrano	BRA_AUT/@@BRA_AUT.ID_BRA_ AUT urlify@@	corago:C 18_Autho r_Role	Responsa bile	RESPONS/@@RESPONS.COD_RE SP urlify@@	crm:E39_Acto r	corago:CP3_ carried_out_ actor
TitoloBr ano	TITOLI/@@TITOLI.TABELLA@@& @@TITOLI.NOME_CAMPO@@& @@TITOLI.COD_BRANO urlify@ @	crm:E35_ Title	Lingua	TADE/LIN/@@TADE.ID_TADE ur lify@@	crm:E56_Lang uage	crm:P72_has _language
TitoloBr anoAltr o	TITOLI/BRA_TIT/@@BRA_TIT.CO D_BRANO@@/@@BRA_TIT.NPR O urlify@@	crm:E35_ Title	TitoloBra noOrdinal e	TITOLI/BRA_TIT/@@BRA_TIT.NP RO urlify@@	crm:E60_Num ber	corago:CPAN 1_priority
TitoloBr anoAltr o	TITOLI/BRA_TIT/@@BRA_TIT.CO D_BRANO@@/@@BRA_TIT.NPR O urlify@@	crm:E35_ Title	BranoTito loTipo	TADE/TIT&@@TADE.ID_TADE ur lify@@	crm:E55_Type	crm:P2_has_ type
TitoloBr anoAltr o	TITOLI/BRA_TIT/@@BRA_TIT.CO D_BRANO@@/@@BRA_TIT.NPR O urlify@@	crm:E35_ Title	Lingua	TADE/LIN/@@TADE.ID_TADE ur lify@@	crm:E56_Lang uage	crm:P72_has _language
TitoloPi anoBran o	TITOLI/SPET_BRA_TIT/@@SPET_ BRA_TIT.ID_SPE_T_BRA urlify@@ /@@SPET_BRA_TIT.NPRO urlify @@	crm:E35_ Title	BranoTito loTipo	TADE/TIT&@@TADE.ID_TADE ur lify@@	crm:E55_Type	crm:P2_has_ type

TitoloPianoBranco	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.ID_SPET_BRA urlify@@/@@SPET_BRA_TIT.NPRO urlify@@	crm:E35_Title	TitoloPianoBrancoOrdinale	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.NPRO urlify@@	crm:E60_Number	corago:CPAN1_priority
TitoloPianoBranco	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.ID_SPET_BRA urlify@@/@@SPET_BRA_TIT.NPRO urlify@@	crm:E35_Title	Lingua	TADE/LIN/@@TADE.ID_TADE urlify@@	crm:E56_Language	crm:P72_has_language

A.3. Mapping degli Spettacoli

Il diagramma rappresenta l'area del modello relativa agli Spettacoli, la relativa relazione con i Brani e le responsabilità autoriali. Completano l'area le relazioni con spazio (collocazione geografica) e tempo (essendo la classe frbroo:F31 Performance una specializzazione di crm:E5 Event)¹⁸³.

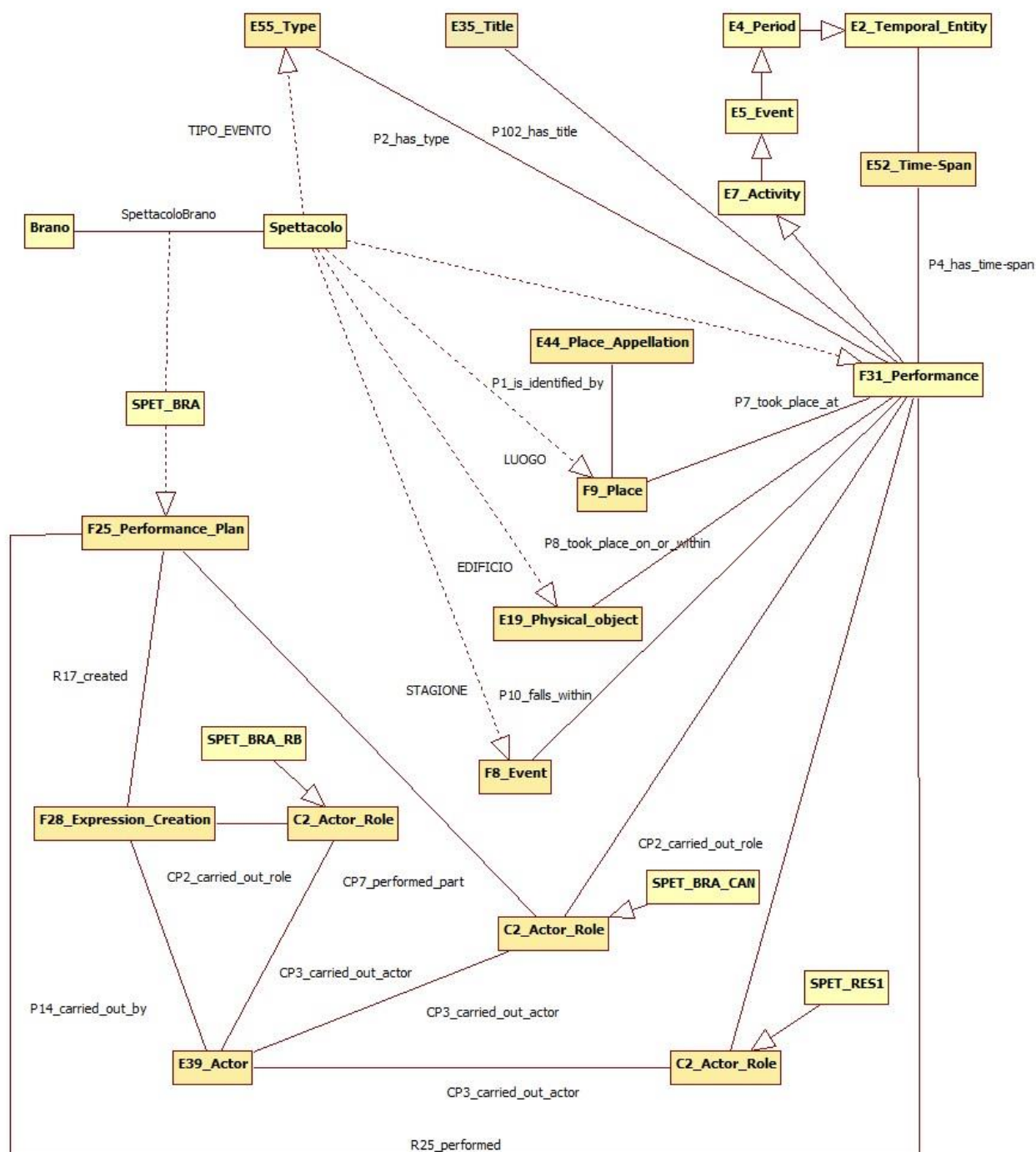


Diagramma A.3: Mapping dell'entità Spettacolo

¹⁸³ In relazione alla normalizzazione delle informazioni relative alla collocazione nello spazio e tempo delle diverse entità si rinvia al capitolo 3.

Se per l'entità Spettacolo si è potuto procedere ad un *mapping* diretto con la classe frbroo:F31_Performance, più complessa è risultata invece la resa della relazione tra Spettacolo e Brani. Si è deciso di adottare la classe frbroo:F25_Performance_Plan come *expression* dei brani che costituiscono il programma dell'evento, relazione che, in Corago, viene tracciata nella tabella SPET_BRA. Sulle istanze di quest'ultima vengono attestate tre macro-tipologie di responsabilità ciascuna con i relativi ruoli: autori, cantanti e responsabili allestimento (ciascuna censita in una tabella dedicata). Nel caso delle responsabilità degli autori dei brani, essendo frbroo:F25_Performance_Plan una sottoclasse di frbroo:F2_Expression richiede una istanza di frbroo:F28_Expression_Creation cui associare il ruolo del relativo crm:E39_Actor¹⁸⁴.

Per la gestione della tipologia di spettacolo si è utilizzato direttamente crm:E55_Type: senza specializzare in *Corago* ed utilizzando TADE come vocabolario.

I titoli sono derivati dalla tabella SPET_BRA_TIT ed attestati sulla classe crm:E35_Title attraverso la proprietà crm:P1_is_identified_by.

Per la gestione della collocazione spaziale dell'evento, l'attributo Luogo viene definito come frbroo:F9_Place¹⁸⁵ mentre l'edificio come crm:E19_Physical_object.

A.3.1. Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
Spettacolo	frbroo:F31_Performance	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	rdfs:label	SPETTACOLI.TIT_NO R
Spettacolo	frbroo:F31_Performance	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	<http://corago.unibo.it/sm/CDP5_season_sequence_order>	SPETTACOLI.NUM_S PET_STA
Spettacolo	frbroo:F31_Performance	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	<http://corago.unibo.it/sm/CDP6_performance_title_in_source>	SPETTACOLI.TIT_PRO PRIO

A.3.2. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	RinvioBibli SpettacoloNota	LEGAMI/RIN_SPE/NOTE/@@RIN_BIB.ID_RINV urlify@@	crm:E62_String	corago:CPBC37_performanc e_ref_note
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	Stagione	STAGIONI/@@STAGIONI.ID_STAG urlify@@	corago:C21_Season	corago:CPBC47_performanc e_season
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	TipoEvento	TADE/EVN&@@TADE.ID_TADE urlify@@	corago:C14_Event_Type	corago:CPBC10_type

¹⁸⁴ Questa è una caratteristica peculiare di FRBROO che prevede la mediazione di una classe Evento tra l'Attore e l'Oggetto prodotto.

¹⁸⁵ A sua volta identificato da crm:E44_Place_Appellation cui in futuro potranno essere associati i riferimenti per la geolocalizzazione.

Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	RinvioBiblioSpettacolo	LEGAMI/RIN_SPE/@@RIN_BIB.ID_RINV urlify@@	corago:C5_Link	corago:CPBC16_refers_to
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	PianoBrano	SPET_BRA/@@SPET_BRA.ID_SPET_BRA urlify@@	frbroo:F25_Performance_Plan	corago:CPBC2_performed
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	LuogoSpettacolo	SPETTACOLI/LUOGO/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F9_Place	corago:CPBC3_took_place_at
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	EdificioSpettacolo	SPETTACOLI/EDIFICIO/@@SPETTACOLI.EDIFICIO urlify@@	crm:E19_Physical_object	corago:CPBC3_took_place_at
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	Personaggio	PERSONAG/@@PERSONAG.COD_PERSONER urlify@@	frbroo:F38_Character	corago:CPBC5_character_performance
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	NoteSpettacolo	SPETTACOLI/NOTE/@@SPETTACOLI.ID_SPET urlify@@	crm:E62_String	corago:CPBC6_note
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CPBC7_performance_actor
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	DataSpettacolo	SPETTACOLI/DATA/@@SPETTACOLI.ID_SPET urlify@@	crm:E52_Time-Span	crm:P4_has_time-span
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	EdificioSpettacolo	SPETTACOLI/EDIFICIO/@@SPETTACOLI.EDIFICIO urlify@@	crm:E19_Physical_object	crm:P8_took_place_on_or_within
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	LuogoSpettacolo	SPETTACOLI/LUOGO/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F9_Place	crm:P7_took_place_at
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	NoteSpettacolo	SPETTACOLI/NOTE/@@SPETTACOLI.ID_SPET urlify@@	crm:E62_String	crm:P3_has_note
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	DataSpettacolo	SPETTACOLI/DATA/@@SPETTACOLI.ID_SPET urlify@@	crm:E52_Time-Span	corago:CPBC28_performance_date
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CPBC24_performance_author
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CPBC1_work_author
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	TipoEvento	TADE/EVN&@@TADE.ID_TADE urlify@@	corago:C14_Event_Type	crm:P2_has_type

Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	TitoloSpettacolo	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NOME_CAMPO@@&@@TITOLI.ID_SPET urlify@@	crm:E35_Title	crm:P102_has_title
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	TitoloSpettacolo	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NOME_CAMPO@@&@@TITOLI.ID_SPET urlify@@	crm:E35_Title	corago:CPBC20_title
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	PianoBrano	SPET_BRA/@@SPET_BRA.ID_SPET_BRA urlify@@	frbroo:F25_Performance_Plan	frbroo:R25_performed
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	ProductionRelease	TADE/TAL&@@TADE.ID_TADE urlify@@	corago:C9_Production_Release	crm:P2_has_type
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	TipoOccasione	TADE/OCC&@@TADE.ID_TADE urlify@@	corago:C13_Occasion_Type	crm:P2_has_type
Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	TipoProduzione	TADE/TPP&@@TADE.ID_TADE urlify@@	corago:C12_Production_Type	crm:P2_has_type
Titolo Spettacolo	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NOME_CAMPO@@&@@TITOLI.ID_SPET urlify@@	crm:E35_Title	Lingua	TADE/LIN/@@TADE.ID_TADE urlify@@	crm:E56_Language	crm:P72_has_language
Ruolo Spettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	PriorityRuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@/@@SPET_BRA_CAN.NPRO urlify@@	crm:E60_Number	corago:CPAN1_priority
Ruolo Spettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	RuoloSpettacoloAutNote	SPET_BRA_CAN/NOME_OR/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	crm:E62_String	crm:P3_has_note
Ruolo Spettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	RuoloSpettacoloPerNote	SPET_BRA_CAN/PER_NOME/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	crm:E62_String	crm:P3_has_note
Ruolo Spettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	RuoloSpettacoloQualNote	SPET_BRA_CAN/PER_QUAL/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	crm:E62_String	crm:P3_has_note
Ruolo Spettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	RuoloSpettacoloRuolNote	SPET_BRA_CAN/RUOLO_STO/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	crm:E62_String	crm:P3_has_note
Ruolo Spettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	PianoBrano	SPET_BRA/@@SPET_BRA.ID_SPET_BRA urlify@@	frbroo:F25_Performance_Plan	corago:CP7_performed_part
Ruolo Spettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CP3_carried_out_actor
Ruolo Spettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	Personaggio	PERSONAG/@@PERSONAG.COD_PERSONER urlify@@	frbroo:F38_Character	corago:CP8_performed_character

Ruolo Spettacolo1	SPET_RES1/@@SPET_RES1.ID_SPET_RES1 urlify@@	corago:C20_Contributor_Role	RuoloSpettacolo1Note	SPET_RES1/NOTE/@@SPET_RES1.ID_SPET_RES1 urlify@@	crm:E62_String	crm:P3_has_note
Ruolo Spettacolo1	SPET_RES1/@@SPET_RES1.ID_SPET_RES1 urlify@@	corago:C20_Contributor_Role	RuoloSpettacolo1TipoRuolo	TADE/RUO&@@TADE.ID_TADE urlify@@	corago:C6_Performer_Role	crm:P2_has_type
Ruolo Spettacolo1	SPET_RES1/@@SPET_RES1.ID_SPET_RES1 urlify@@	corago:C20_Contributor_Role	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CP3_carried_out_actor
Luogo Spettacolo	SPETTACOLI/LUOGO/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F9_Place	LuogoSpettacoloPA	LUOGO/@@SPETTACOLI.LUOGO urlify@@	crm:E44_Place_Appellation	crm:P1_is_identified_by

A.4.Mapping dei Personaggi

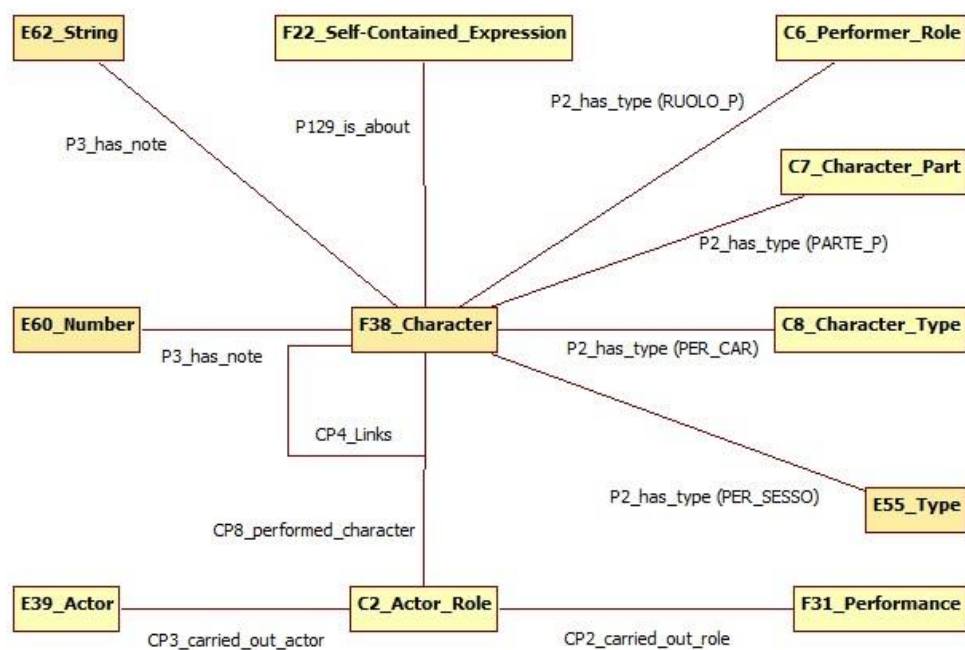


Diagramma A.4: Mapping dell'entità Personaggio

La sezione dedicata alla rappresentazione dei Personaggi è quella in cui sono state introdotte il maggior numero di specializzazioni attraverso il Corago SM. Ciò si è reso necessario per esprimere in modo esaustivo la specificità dei contenuti previste dal modello di origine. Sono state definite classi specifiche per il tipo di personaggio (`corago:C8_Character_Type`), per il ruolo (`corago:C6_Performer_Role`) e la relativa parte (`corago:C7_Character_Part`); tutti concetti non esprimibili in modo esplicito attraverso le due ontologie di riferimento. Per rappresentare la relazione tra interprete e personaggio è stata introdotta la proprietà `corago:CP8_performed_character` che lega l'istanza di `corago:C2_Actor_Role` che specifica il ruolo del `crm:E39_Actor` all'interno della `frbroo:F31_Performance` al `frbroo:F30_Character` che viene interpretato¹⁸⁶. Nel diagramma la proprietà `crm:P129_is_about` che collega la classe `frbroo:F38_Character` alla `frbroo:F22_Self-Contained_Expression` rappresenta l'attestazione di un personaggio al testo utilizzato come fonte. Nel diagramma la classe `frbroo:F22_Self-Contained_Expression` è riportata in quanto *ancestor* comune alle sottoclassi delle istanze cui l'istanza di personaggio è effettivamente associata¹⁸⁷.

A.4.1. Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
Personaggio	frbroo:F38_Character	PERSONAG/@@PERSONAG.COD_PER urlify@@	rdfs:label	PERSONAG.PERSONAG
PersonaggioQualifica	crm:E62_String	PERSONAG/QUAL/@@PERSONAG.QUALIFICA_PER urlify@@	rdfs:label	PERSONAG.QUALIFICA_PER
PersonaggioOrdinale	crm:E60_Number	PERSONAG/PROG/@@PERSONAG.COD_PER urlify@@/@@PERSONAG.PROG_P@@	rdfs:label	PERSONAG.PROG_P

A.4.2. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
RuoloResponsabileEspressione Personaggio	SWT_DOC_BRA_CAN/@@SWT_DOC_BRA_CAN.ID_DOCBRA_CAN urlify@@	corago:C6_Performer_Role	Personaggio	PERSONAG/@@PERSONAG.COD_PER urlify@@	frbroo:F38_Character	corago:CP8_performed_character
RuoloResponsabileEspressione Personaggio	SWT_DOC_BRA_CAN/@@SWT_DOC_BRA_CAN.ID_DOCBRA_CAN urlify@@	corago:C6_Performer_Role	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CP3_carried_out_actor
RuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	PriorityRuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@/@@SPET_BRA_CAN.NPRO urlify@@	crm:E60_Number	corago:CPAN1_priority
RuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	RuoloSpettacoloAutNote	SPET_BRA_CAN/NOME_OR/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	crm:E62_String	crm:P3_has_note

¹⁸⁶ La necessità di poter definire relazioni ternarie tra Attori, Eventi e altre classi è stato uno dei requisiti che ha determinato l'introduzione della classe `corago:C2_Actor_Role`.

¹⁸⁷ Si faccia riferimento al criterio di gestione della specializzazione del Brano in funzione del tipo descritto in A.2.

RuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	RuoloSpettacoloPe rNote	SPET_BRA_CAN/PER_NOME/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	crm:E62_String	crm:P3_has_note
RuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	RuoloSpettacoloQualNote	SPET_BRA_CAN/PER_QUAL/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	crm:E62_String	crm:P3_has_note
RuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	RuoloSpettacoloRuolNote	SPET_BRA_CAN/RUOLO_STO/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	crm:E62_String	crm:P3_has_note
RuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	PianoBrano	SPET_BRA/@@SPET_BRA.ID_SPET_BRA urlify@@	frbroo:F25_Performance_Plan	corago:CP7_performed_part
RuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CP3_carried_out_actor
RuoloSpettacolo	SPET_BRA_CAN/@@SPET_BRA_CAN.ID_BRA_CAN urlify@@	corago:C6_Performer_Role	Personaggio	PERSONAG/@@PERSONAG.COD_PER urlify@@	frbroo:F38_Character	corago:CP8_performed_character
RuoloSpettacolo1	SPET_RES1/@@SPET_RES1.ID_SPET_RES1 urlify@@	corago:C2_O_Contributor_Role	RuoloSpettacolo1Note	SPET_RES1/NOTE/@@SPET_RES1.ID_SPET_RES1 urlify@@	crm:E62_String	crm:P3_has_note
RuoloSpettacolo1	SPET_RES1/@@SPET_RES1.ID_SPET_RES1 urlify@@	corago:C2_O_Contributor_Role	RuoloSpettacolo1TipoRuolo	TADE/RUO&@@TADE.ID_TADE urlify@@	corago:C6_Performer_Role	crm:P2_has_type
RuoloSpettacolo1	SPET_RES1/@@SPET_RES1.ID_SPET_RES1 urlify@@	corago:C2_O_Contributor_Role	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CP3_carried_out_actor
Personaggio	PERSONAG/@@PERSONAG.COD_PER urlify@@	frbroo:F38_Character	EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	corago:CP10_documented_in
Personaggio	PERSONAG/@@PERSONAG.COD_PER urlify@@	frbroo:F38_Character	EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	crm:P67_refers_to
Personaggio	PERSONAG/@@PERSONAG.COD_PER urlify@@	frbroo:F38_Character	NotePersonaggio	PERSONAG/NOTE/@@PERSONAG.NOTE_PER urlify@@	crm:E62_String	crm:P3_has_note
Personaggio	PERSONAG/@@PERSONAG.COD_PER urlify@@	frbroo:F38_Character	PersonaggioOrdinale	PERSONAG/PROG/@@PERSONAG.COD_PER urlify@@/@@PERSONAG.PROG_P@@	crm:E60_Number	corago:CPAN1_priority
Personaggio	PERSONAG/@@PERSONAG.COD_PER urlify@@	frbroo:F38_Character	PersonaggioQualifica	PERSONAG/QUAL/@@PERSONAG.QUALIFICA_PER urlify@@	crm:E62_String	crm:P3_has_note
Personaggio	PERSONAG/@@PERSONAG.COD_PER urlify@@	frbroo:F38_Character	RuoloInterprete	TADE/RUO&@@TADE.ID_TADE urlify@@	corago:C6_Performer_Role	crm:P2_has_type

Personaggio	PERSONAG/@@PERSONA G.COD_PER urlify@@	frbroo:F38 _Characte r	SessoPers onaggio	TADE/RUO&@@TADE.ID_TADE url ify@@	crm:E55_Ty pe	crm:P2_has _type
Personaggio	PERSONAG/@@PERSONA G.COD_PER urlify@@	frbroo:F38 _Characte r	TypePers onaggio	TADE/CRT&@@TADE.ID_TADE urlif y@@	corago:C8_C haracter_Ty pe	crm:P2_has _type

A.5.Mapping dei Titoli

Lo schema di associazione prevista dal modello di origine per i titoli evidenzia come particolarità la cardinalità 1-*n* delle relazioni tra le singole istanze di frbroo:F1_Work ed i relativi titoli censiti nella tabella associativa BRA_TIT. Nella tabella, assieme alla relazione, sono presenti alcuni attributi che sono stati mappati sulle relative classi del modello di destinazione (vedi A.5.1). Stessa considerazione vale per le classi frbroo:F25_Performance_Plan e frbroo:F31_Performance che sono state associate attraverso la tabella SPET_BRA_TIT.

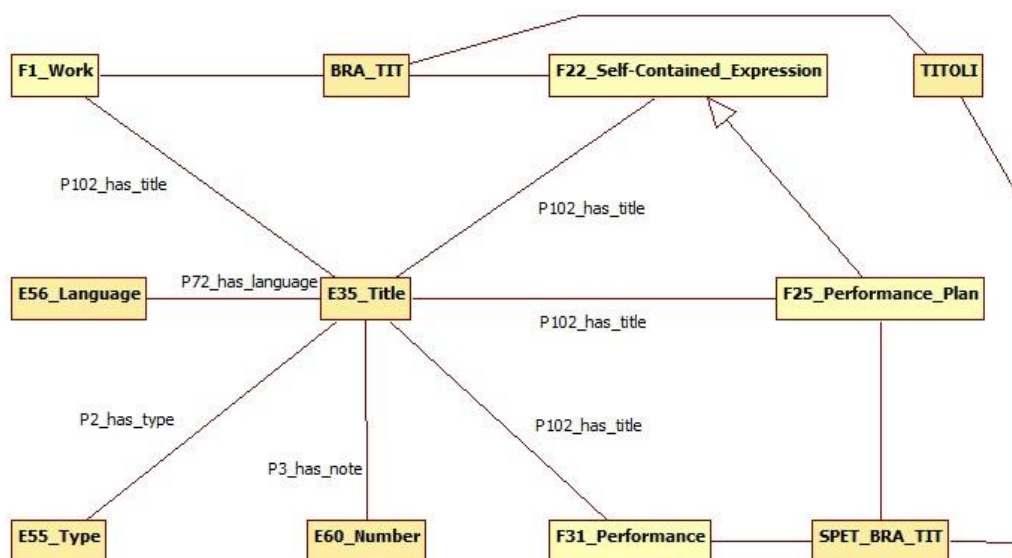


Diagramma A.5: Mapping dei Titoli

A.5.1. Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
TitoloBrano	crm:E35_Tit le	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NOME_CAMPO@@&@@TI TOLI.COD_BRANO urlify@@	rdfs:la bel	TITOLI.TITOLO
TitoloBranoAltro	crm:E35_Tit le	TITOLI/BRA_TIT/@@BRA_TIT.COD_BRANO@@/@@BRA_TIT.NPRO urlif y@@	rdfs:la bel	BRA_TIT.TIT_NOR
TitoloDocumento	crm:E35_Tit le	TITOLI/DOC_BRA_TIT/@@DOC_BRA_TIT.ID_DOC_BRA urlify@@/@@D OC_BRA_TIT.NPRO urlify@@	rdfs:la bel	DOC_BRA_TIT.TIT_ NOR

TitoloDocumento Dir	crm:E35_Tit le	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NOME_CAMPO@@&@@TI TOLI.ID_DOC urlify@@	rdfs:la bel	TITOLI.TITOLO
TitoloLibBibliograf ia	crm:E35_Tit le	BIBLIOGRAFIA/TIT_LIB/@@BIBLIOGRAFIA.ID_BIB urlify@@	rdfs:la bel	BIBLIOGRAFIA.TITO LO_LIB
TitoloPianoBrano	crm:E35_Tit le	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.ID_SPET_BRA urlify@@/@@S PET_BRA_TIT.NPRO urlify@@	rdfs:la bel	SPET_BRA_TIT.TIT_ NOR
TitoloSagBibliogra fia	crm:E35_Tit le	BIBLIOGRAFIA/TIT_SAG/@@BIBLIOGRAFIA.ID_BIB urlify@@	rdfs:la bel	BIBLIOGRAFIA.TIT OLO_SAG
TitoloSezione	crm:E35_Tit le	SEZ/TITOLI/@@SEZ1.ID_SEZ1 urlify@@	rdfs:la bel	SEZ1.TIT_SEZ1
TitoloSpettacolo	crm:E35_Tit le	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NOME_CAMPO@@&@@TI TOLI.ID_SPET urlify@@	rdfs:la bel	TITOLI.TITOLO
TitoloBranoOrdina le	crm:E60_Nu mber	TITOLI/BRA_TIT/@@BRA_TIT.NPRO urlify@@	rdfs:la bel	BRA_TIT.NPRO
TitoloDocumento Ordinale	crm:E60_Nu mber	TITOLI/DOC_BRA_TIT/@@DOC_BRA_TIT.NPRO urlify@@	rdfs:la bel	DOC_BRA_TIT.NPR O
TitoloPianoBrano Ordinale	crm:E60_Nu mber	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.NPRO urlify@@	rdfs:la bel	SPET_BRA_TIT.NPR O

A.5.2. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associat a	Relazione
TitoloBra no	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NO ME_CAMPO@@&@@TITOLI.COD_BRANO urlif y@@	crm:E 35_Tit le	Lingua	TADE/LIN/@@TADE.ID_TA DE urlify@@	crm:E56 _Langua ge	crm:P72_h as_langua ge
TitoloBra noAltro	TITOLI/BRA_TIT/@@BRA_TIT.COD_BRANO@@/ @@BRA_TIT.NPRO urlify@@	crm:E 35_Tit le	TitoloBrano Ordinale	TITOLI/BRA_TIT/@@BRA_T IT.NPRO urlify@@	crm:E60 _Numbe r	corago:CP AN1_piori ty
TitoloBra noAltro	TITOLI/BRA_TIT/@@BRA_TIT.COD_BRANO@@/ @@BRA_TIT.NPRO urlify@@	crm:E 35_Tit le	BranoTitolo Tipo	TADE/TIT&@@TADE.ID_TA DE urlify@@	crm:E55 _Type	crm:P2_ha s_type
TitoloBra noAltro	TITOLI/BRA_TIT/@@BRA_TIT.COD_BRANO@@/ @@BRA_TIT.NPRO urlify@@	crm:E 35_Tit le	Lingua	TADE/LIN/@@TADE.ID_TA DE urlify@@	crm:E56 _Langua ge	crm:P72_h as_langua ge
TitoloDoc umento	TITOLI/DOC_BRA_TIT/@@DOC_BRA_TIT.ID_DO C_BRA urlify@@/@@DOC_BRA_TIT.NPRO urlif y@@	crm:E 35_Tit le	TitoloDocu mentoOrdin ale	TITOLI/DOC_BRA_TIT/@@ DOC_BRA_TIT.NPRO urlify @@	crm:E60 _Numbe r	corago:CP AN1_piori ty
TitoloDoc umento	TITOLI/DOC_BRA_TIT/@@DOC_BRA_TIT.ID_DO C_BRA urlify@@/@@DOC_BRA_TIT.NPRO urlif y@@	crm:E 35_Tit le	BranoTitolo Tipo	TADE/TIT&@@TADE.ID_TA DE urlify@@	crm:E55 _Type	crm:P2_ha s_type

TitoloDocumento	TITOLI/DOC_BRA_TIT/@@DOC_BRA_TIT.ID_DOC_BRA urlify@@/@@DOC_BRA_TIT.NPRO urlify@@	crm:E35_Title	Lingua	TADE/LIN/@@TADE.ID_TADE urlify@@	crm:E56_Language	crm:P72_has_language
TitoloDocumentoDir	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NO_ME_CAMPO@@&@@TITOLI.ID_DOC urlify@@	crm:E35_Title	Lingua	TADE/LIN/@@TADE.ID_TADE urlify@@	crm:E56_Language	crm:P72_has_language
TitoloPianoBrano	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.ID_SPET_BRA urlify@@/@@SPET_BRA_TIT.NPRO urlify@@	crm:E35_Title	BranoTitoloTipo	TADE/TIT&@@TADE.ID_TADE urlify@@	crm:E55_Type	crm:P2_has_type
TitoloPianoBrano	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.ID_SPET_BRA urlify@@/@@SPET_BRA_TIT.NPRO urlify@@	crm:E35_Title	TitoloPianoBranoOrdinale	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.NPRO urlify@@	crm:E60_Number	corago:CPAN1_priority
TitoloPianoBrano	TITOLI/SPET_BRA_TIT/@@SPET_BRA_TIT.ID_SPET_BRA urlify@@/@@SPET_BRA_TIT.NPRO urlify@@	crm:E35_Title	Lingua	TADE/LIN/@@TADE.ID_TADE urlify@@	crm:E56_Language	crm:P72_has_language
TitoloSpettacolo	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NO_ME_CAMPO@@&@@TITOLI.ID_SPET urlify@@	crm:E35_Title	Lingua	TADE/LIN/@@TADE.ID_TADE urlify@@	crm:E56_Language	crm:P72_has_language

A.6.Mapping dei Legami

Il modello Corago presenta una considerevole analiticità nella definizione dei tipi di legame che mettono in relazione le diverse entità del modello concettuale. Per questa ragione si è deciso di introdurre nel Corago SM una classe che esprimesse il concetto di legame per rappresentare le diverse tipologie previste dal modello originale mediante attributi descrittivi specifici. Il diagramma rappresenta il modello *target* dell'estrazione del contenuto della tabella LEGAMI che è stato applicato in tutti i casi in cui fosse presente in *Corago* un legame tra istanze omogenee e non.

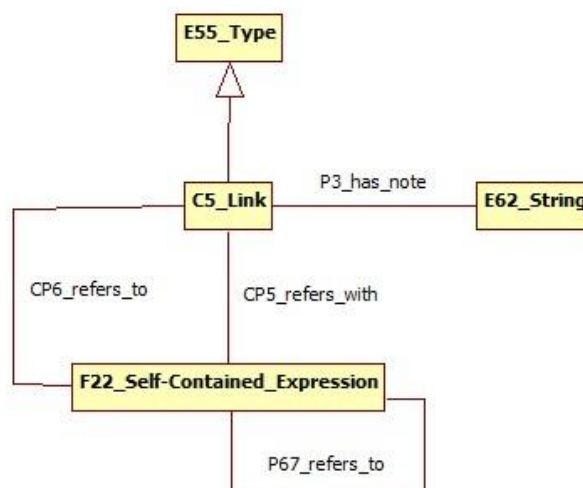


Diagramma A.6: Mapping dei Legami

A.6.1. Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
Legame	corago:C5_Link	LEGAMI/@@TADE.TADETIP@@/@@LEGAMI.COD_BRANO@@-@@LEGAMI.COD_BRANOL urlify@@	rdfs:label	LEGAMI.TM
Legame	corago:C5_Link	LEGAMI/@@TADE.TADETIP@@/@@LEGAMI.COD_BRANO@@-@@LEGAMI.COD_BRANOL urlify@@	rdfs:label	TADE.TADEDA1
LegameDocumento	corago:C5_Link	LEGAMI/@@TADE.TADETIP@@/@@SWT_LEGAMI_DOC.ID_DOC@@-@@SWT_LEGAMI_DOC.ID_DOC_PADRE urlify@@	rdfs:label	TADE.TADEDA1
RinvioBiblioSpettacolo	corago:C5_Link	LEGAMI/RIN_SPE/@@RIN_BIB.ID_RINV urlify@@	rdfs:label	RIN_BIB.RINVIO
RinvioBiblioID	crm:E42_Identifier	LEGAMI/RIN/ID/@@SWT_DOC_BIB.ID_RINV urlify@@	rdfs:label	SWT_DOC_BIB.RIN_ID
RinvioBiblioURL	crm:E42_Identifier	LEGAMI/RIN/ID/URL/@@SWT_DOC_BIB.ID_RINV urlify@@	rdfs:label	BIBLIOGRAFIA.BIB_URL1 & SWT_DOC_BIB.RIN_ID & BIBLIOGRAFIA.BIB_URL2
RinvioBiblioBranoOrdinale	crm:E60_Number	LEGAMI/RIN_BRA/ORD/@@SWT_BRANI_BIB.ID_RINV urlify@@	rdfs:label	SWT_BRANI_BIB.NPRO
RinvioBiblioOrdinale	crm:E60_Number	LEGAMI/RIN/ORD/@@SWT_DOC_BIB.ID_RINV urlify@@	rdfs:label	SWT_DOC_BIB.NPRO
RinvioBiblioSpettacoloOrdinale	crm:E60_Number	LEGAMI/RIN_SPE/ORD/@@RIN_BIB.ID_RINV urlify@@	rdfs:label	RIN_BIB.NPRO
RinvioBiblioBranoNota	crm:E62_String	LEGAMI/RIN_BRA/NOTE/@@SWT_BRANI_BIB.ID_RINV urlify@@	rdfs:label	SWT_BRANI_BIB.RIN_NOTE
RinvioBiblioNota	crm:E62_String	LEGAMI/RIN/NOTE/@@SWT_DOC_BIB.ID_RINV urlify@@	rdfs:label	SWT_DOC_BIB.RIN_NOTE
RinvioBiblioSpettacoloNota	crm:E62_String	LEGAMI/RIN_SPE/NOTE/@@RIN_BIB.ID_RINV urlify@@	rdfs:label	RIN_BIB.RIN_NOTE

A.6.2. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
Legame	LEGAMI/@@TADE.TADETIP@@/@@LEGAMI.COD_BRANO@@-@@LEGAMI.COD_BRANOL urlify@@	corago:C5_Link	NoteLegame	LEGAMI/NOTE/@@LEGAMI.NOTE_LEG urlify@@	crm:E62_String	crm:P3_has_note
LegameDocumento	LEGAMI/@@TADE.TADETIP@@/@@SWT_LEGAMI_DOC.ID_DOC@@-@@	corago:C5_Link	Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	corago:CP6_refers_to

	@@SWT_LEGAMI_DOC.ID_DOC_PADRE urlify@@					
RinvioBiblio	LEGAMI/RIN/@@SWT_DOC_BIB.ID_RINV urlify@@	corago:C19_Bibliographic_Reference	RinvioBibliolD	LEGAMI/RIN/ID/@@SWT_DOC_BIB.ID_RINV urlify@@	crm:E42_Identifier	crm:P1_is_identified_by
RinvioBiblio	LEGAMI/RIN/@@SWT_DOC_BIB.ID_RINV urlify@@	corago:C19_Bibliographic_Reference	RinvioBiblioNote	LEGAMI/RIN/NOTE/@@SWT_DOC_BIB.ID_RINV urlify@@	crm:E62_String	crm:P3_has_note
RinvioBiblio	LEGAMI/RIN/@@SWT_DOC_BIB.ID_RINV urlify@@	corago:C19_Bibliographic_Reference	RinvioBiblioOrdinale	LEGAMI/RIN/ORD/@@SWT_DOC_BIB.ID_RINV urlify@@	crm:E60_Number	corago:CPAN1_priority
RinvioBiblio	LEGAMI/RIN/@@SWT_DOC_BIB.ID_RINV urlify@@	corago:C19_Bibliographic_Reference	RinvioBiblioURL	LEGAMI/RIN/ID/URL/@@SWT_DOC_BIB.ID_RINV urlify@@	crm:E42_Identifier	corago:CP15_has_URL
RinvioBiblio	LEGAMI/RIN/@@SWT_DOC_BIB.ID_RINV urlify@@	corago:C19_Bibliographic_Reference	Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	corago:CP6_refers_to
RinvioBiblioBrano	LEGAMI/RIN_BRA/@@SWT_BRANI_BIB.ID_RINV urlify@@	corago:C19_Bibliographic_Reference	RinvioBiblioBranoNota	LEGAMI/RIN_BRA/NOTE/@@SWT_BRANI_BIB.ID_RINV urlify@@	crm:E62_String	crm:P3_has_note
RinvioBiblioBrano	LEGAMI/RIN_BRA/@@SWT_BRANI_BIB.ID_RINV urlify@@	corago:C19_Bibliographic_Reference	RinvioBiblioBranoOrdinale	LEGAMI/RIN_BRA/ORD/@@SWT_BRANI_BIB.ID_RINV urlify@@	crm:E60_Number	corago:CPAN1_priority
RinvioBiblioSpettacolo	LEGAMI/RIN_SPE/@@RIN_BIB.ID_RINV urlify@@	corago:C5_Link	RinvioBiblioSpettacoloNota	LEGAMI/RIN_SPE/NOTE/@@RIN_BIB.ID_RINV urlify@@	crm:E62_String	crm:P3_has_note
RinvioBiblioSpettacolo	LEGAMI/RIN_SPE/@@RIN_BIB.ID_RINV urlify@@	corago:C5_Link	RinvioBiblioSpettacoloOrdinale	LEGAMI/RIN_SPE/ORD/@@RIN_BIB.ID_RINV urlify@@	crm:E60_Number	corago:CPAN1_priority
RinvioBiblioSpettacolo	LEGAMI/RIN_SPE/@@RIN_BIB.ID_RINV urlify@@	corago:C5_Link	Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@@	frbroo:F31_Performance	corago:CP6_refers_to

A.7.Mapping delle Sezioni

L'articolazione delle opere in atti, quadri scenici o parti è definita in una tabella specifica denominata: SEZ1. Per ciascuna sezione viene riportato titolo, numero ed articolazione. Per la trasposizione è stata seguita l'indicazione suggerita dalle stesse linee guida per l'implementazione di FRBRoo che prevedono di trattare le sezioni come frbroo:F1_Work associate al Brano di riferimento attraverso una relazione R10_has_member. Il diagramma A.7 rappresenta l'articolazione della Sezione e dei suoi attributi.

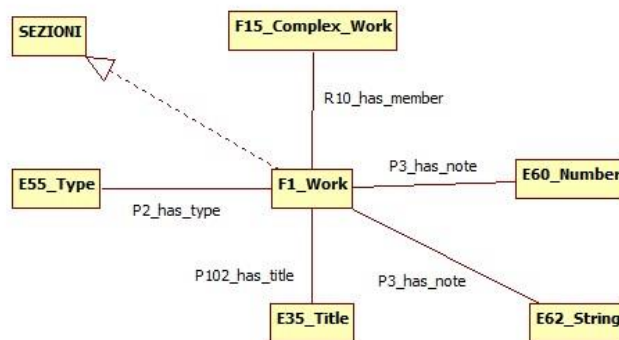


Diagramma A.7: Mapping delle Sezioni

A.7.1. Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
SezioneNumero	crm:E62_String	SEZ/NSEZ/@@SEZ1.ID_SEZ1 urlify@@/@@SEZ1.NSEZ1 urlify@@	rdfs:label	SEZ1.NSEZ1
SezioneTesto	crm:E62_String	SEZ/TESTO/@@SEZ1.ID_SEZ1 urlify@@	rdfs:label	SEZ1.TESTO
PrioritySezione	crm:E60_Number	SEZ/NPRO/@@SEZ1.ID_SEZ1 urlify@@/@@SEZ1.NSEZ1 urlify@@	rdfs:label	SEZ1.NSEZ1
TitoloSezione	crm:E35_Title	SEZ/TITOLI/@@SEZ1.ID_SEZ1 urlify@@	rdfs:label	SEZ1.TIT_SEZ1

A.7.2. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
Sezione	SEZ/@@SEZ1.ID_SEZ1 urlify@@	frbroo:F23_Expression_Fragment	PrioritySezione	SEZ/NPRO/@@SEZ1.ID_SEZ1 urlify@@/@@SEZ1.NSEZ1 urlify@@	crm:E60_Number	corago:CPAN1_priority
Sezione	SEZ/@@SEZ1.ID_SEZ1 urlify@@	frbroo:F23_Expression_Fragment	SezioneNumero	SEZ/NSEZ/@@SEZ1.ID_SEZ1 urlify@@/@@SEZ1.NSEZ1 urlify@@	crm:E62_String	crm:P3_has_note
Sezione	SEZ/@@SEZ1.ID_SEZ1 urlify@@	frbroo:F23_Expression_Fragment	SezioneOrdinale	SEZ/SCENE/@@SEZ1.ID_SEZ1 urlify@@/@@SEZ1.SCENA urlify@@	crm:E60_Number	crm:P3_has_note
Sezione	SEZ/@@SEZ1.ID_SEZ1 urlify@@	frbroo:F23_Expression_Fragment	SezioneTesto	SEZ/TESTO/@@SEZ1.ID_SEZ1 urlify@@	crm:E62_String	crm:P3_has_note
Sezione	SEZ/@@SEZ1.ID_SEZ1 urlify@@	frbroo:F23_Expression_Fragment	TipoSezione	TADE/SZ1&@@TADE.ID_TADE urlify@@	corago:E55_Type	crm:P2_has_type
Sezione	SEZ/@@SEZ1.ID_SEZ1 urlify@@	frbroo:F23_Expression_Fragment	TitoloSezione	SEZ/TITOLI/@@SEZ1.ID_SEZ1 urlify@@	crm:E35_Title	crm:P102_has_title
Sezione	SEZ/@@SEZ1.ID_SEZ1 urlify@@	frbroo:F23_Expression_Fragment	TitoloSezione	SEZ/TITOLI/@@SEZ1.ID_SEZ1 urlify@@	crm:E35_Title	corago:CPBC20_title

Il diagramma A.8 illustra l'ambito del modello relativo ai Documenti in cui in *Corago* viene raccolta tutta l'informazione relativa alle diverse fonti documentarie.



Se la corrispondenza tra i Documenti in *Corago* e la classe frbroo:F3_Manifestation_Product_Type è risultata abbastanza lineare, per ricostruire il contesto editoriale del documento storico è necessario l'impiego di due classi intermedie: frbroo:F24_Publication_Expression e frbroo:F30_Publication_Event. Sulla prima vengono attestati tutti i riferimenti alle classi frbroo:F1_Work e frbroo:F2_Expression, la seconda invece raccoglie gli attributi propri del processo editoriale.

Anche in questo caso, per l'esposizione delle relazioni tra Documenti, è stata impiegata la classe `corago:C5` [Link per specificarne la tipologia](#). Come nel caso del *mapping* della tabella SPET BRA CAN, anche

per la SWT_DOC_BRA_CAN si è ricorso alla classe corago:C2_Actor_Role per creare la relazione ternaria tra Attore, Spettacolo e *Publication Expression*; rimane da approfondire la necessità di utilizzare la proprietà specializzata corago:CP9_documented_in che permette di riferire la responsabilità rispetto al documento.

A.8.1. Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
DataEdizione	crm:E52_Time-Span	DOCUMENTI/DATA/@@DOCUMENTI.ID_DOC urlify@@	rdfs:label	DOCUMENTI.DOC_DATA
DocumentoDimensione	crm:E54_Dimension	DOCUMENTI/DIM/@@DOCUMENTI.ID_DOC urlify@@	rdfs:label	DOCUMENTI.DOC_DIMENSIONI
DocumentoPaginazione	crm:E62_String	DOCUMENTI/PAG/@@DOCUMENTI.ID_DOC urlify@@	rdfs:label	DOCUMENTI.DOC_PAGINAZIONE
NoteDocumento	crm:E62_String	DOCUMENTI/NOTE/@@DOCUMENTI.ID_DOC urlify@@	rdfs:label	DOCUMENTI.DOC_NOTE

A.8.2. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	EspressioneDocumentoNote	SWT_DOC_BRA/NOTE/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	crm:E62_String	corago:CPBC41_document_work_notes
Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	ExplicitDocumentoStruttura	DOC_STRUTTURA/DS_EXPLICIT/@@DOC_STRUTTURA.ID_DS urlify@@	crm:E33_Linguistic_Object	corago:CPBC38_document_explicit
Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	ItemColloc	ITEM/MCOLLOC/@@SWT_DOC_RISM.ID_MBIBLIO@@	crm:E62_String	corago:CPBC44_document_collocation
Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	ItemPermLink	ITEM/MBIB_ID/@@SWT_DOC_RISM.ID_MBIBLIO@@	crm:E42_Identifier	corago:CPBC45_document_biblio_id
Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	EditoreDocumentoNome	DOCUMENTI/EDITORE/@@DOCUMENTI.DOC_EDITORE urlify@@	crm:E82_ActorAppellation	corago:CPBC21_published_by
Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	LuogoEdizionePA	LUOGO/@@DOCUMENTI.DOC_LUOGO urlify@@	crm:E44_PlaceAppellation	corago:CPBC22_place_of_publication

Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	LuogoStruttu ra	DOC_STRUTTURA/DS_LUOGO/ @@DOC_STRUTTURA.ID_DS ur lify@@	frbroo:F9_Pi ace	corago:CPBC3 9_document_ struct_place
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	NoteDocume nto	DOCUMENTI/NOTE/@@DOCU MENTI.ID_DOC urlify@@	crm:E62_Str ing	corago:CPBC6 _note
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	Noteltem	ITEM/NOTE/@@SWT_DOC_RIS M.MBIB_NOTE urlify@@	crm:E62_Str ing	corago:CPBC4 3_document_ item_notes
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	Responsabile	RESPONS/@@RESPONS.COD_R ESP urlify@@	crm:E39_Ac tor	corago:CPBC2 3_work_auth or_in_source
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	RinvioBiblioN ote	LEGAMI/RIN/NOTE/@@SWT_D OC_BIB.ID_RINV urlify@@	crm:E62_Str ing	corago:CPBC4 2_document_ reference_no tes
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	RismAppellat ion	RISM/APP/@@RISM.DENOMIN AZIONE urlify@@	crm:E82_Ac tor_Appellat ion	corago:CPBC4 6_document_ biblio_denom
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	DocumentoP aginazione	DOCUMENTI/PAG/@@DOCU MENTI.ID_DOC urlify@@	crm:E62_Str ing	corago:CPBC4 8_document_ pag
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	Espressioned ocumentoTip o	SWT_DOC_BRA/FORMA/@@S WT_DOC_BRA.ID_DOC_BRA url ify@@	crm:E55_Ty pe	corago:CPBC4 0_document_ content_genr e
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	LegameDocu mento	LEGAMI/@@TADE.TADETIP/@@ /@@SWT_LEGAMI_DOC.ID_DO C@@- @@SWT_LEGAMI_DOC.ID_DOC _PADRE urlify@@	corago:C5_L ink	corago:CPBC1 6_refers_to
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	RinvioBiblio	LEGAMI/RIN/@@SWT_DOC_BI B.ID_RINV urlify@@	corago:C19_ Bibliographi c_Reference	corago:CPBC1 6_refers_to
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	Espressioned ocumento	SWT_DOC_BRA/@@SWT_DOC_ BRA.ID_DOC_BRA urlify@@	frbroo:F24_ Publication_ Expression	frbroo:CLR6_s hould_carry
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestati on_Product _Type	Spettacolo	SPETTACOLI/@@SPETTACOLI.ID _SPET urlify@@	frbroo:F31_ Performanc e	crm:P67_refe rs_to

Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	DocumentoD imensione	DOCUMENTI/DIM/@@DOCUMENTI.ID_DOC urlify@@	crm:E54_Di mension	crm:P43_has_ dimension
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	StrutturaDoc umento	DOC_STRUTTURA/@@DOC_STRUTTURA.ID_DS urlify@@	frbroo:F23_Expression_Fragment	corago:CPBC25_document_fragment
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	Lingua	TADE/LIN/@@TADE.ID_TADE urlify@@	crm:E56_La nguage	crm:P72_has_ language
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	DocumentoN ote	DOCUMENTI/NOTE/@@DOCUMENTI.ID_DOC urlify@@	crm:E62_Str ing	crm:P3_has_ note
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	LegameDocu mento	LEGAMI/@@TADE.TADETIP@@/@@SWT_LEGAMI_DOC.ID_DOC@@-@@SWT_LEGAMI_DOC.ID_DOC_PADRE urlify@@	corago:C5_L ink	corago:CP5_r efers_with
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	DocumentoP aginazione	DOCUMENTI/PAG/@@DOCUMENTI.ID_DOC urlify@@	crm:E62_Str ing	crm:P3_has_ note
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	TestoDocum entoStruttura	DOC_STRUTTURA/DS_TESTO/@@DOC_STRUTTURA.ID_DS urlify@@	crm:E33_Lin guistic_Obje ct	corago:CPBC26_document_fragment_text
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	TipoEdizione	TADE/EDL/@@TADE.ID_TADE urlify@@	corago:C15_Edition_Type	crm:P2_has_t ype
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	NoteDocume nto	DOCUMENTI/NOTE/@@DOCUMENTI.ID_DOC urlify@@	crm:E62_Str ing	crm:P3_has_ note
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	StrutturaDoc umento	DOC_STRUTTURA/@@DOC_STRUTTURA.ID_DS urlify@@	frbroo:F23_Expression_Fragment	frbroo:R15_h as_fragment
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	TitoloDocum entoDir	TITOLI/@@TITOLI.TABELLA@@&@@TITOLI.NOME_CAMPO@@&@@TITOLI.ID_DOC urlify@@	crm:E35_Titl e	crm:P102_ha s_title
Documen to	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	RuoloRespon sabileDocum ento	SWT_DOC_RESP/@@SWT_DOC_RESP.ID_DOC_RESP@@	corago:C2_A ctor_Role	corago:CP2_c arried_out_ro le

Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	TipoDocumento	TADE/LB1&@@TADE.ID_TADE urlify@@	crm:E55_Type	crm:P2_has_type
EditoreDocumento	DOCUMENTI/@@DOCUMENTI.ID_DOC@@/EDITORE	crm:E39_Actor	EditoreDocumentoNome	DOCUMENTI/EDITORE/@@DOCUMENTI.DOC_EDITORE urlify@@	crm:E82_ActorAppellation	crm:P131_is_identified_by
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	Lingua	TADE/LIN/@@TADE.ID_TADE urlify@@	crm:E56_Language	crm:P72_has_language
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	PriorityEspressioneDocumento	SWT_DOC_BRA/NPRO/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@/@@SWT_DOC_BRA.NPRO@@	crm:E60_Number	corago:CPAN1_priority
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	EspressioneDocumentoNote	SWT_DOC_BRA/NOTE/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	crm:E62_String	crm:P3_has_note
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	EspressioneDocumentoTipo	SWT_DOC_BRA/FORMA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	crm:E55_Type	crm:P2_has_type
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	Genere	TADE/GEN&@@TADE.ID_TADE urlify@@	corago:C3_Genre	crm:P2_has_type
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	ProductionRelease	TADE/TAL&@@TADE.ID_TADE urlify@@	corago:C9_Production_Release	crm:P2_has_type
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	TitoloDocumento	TITOLI/DOC_BRA_TIT/@@DOC_BRA_TIT.ID_DOC_BRA urlify@@/@@DOC_BRA_TIT.NPRO urlify@@	crm:E35_Title	crm:P102_has_title
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	TitoloDocumento	TITOLI/DOC_BRA_TIT/@@DOC_BRA_TIT.ID_DOC_BRA urlify@@/@@DOC_BRA_TIT.NPRO urlify@@	crm:E35_Title	corago:CPBC20_title
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	RuoloResponsabileEspressioneDocumento	SWT_DOC_BRA_RB/@@SWT_DOC_BRA_RB.ID_DOCBRA_RB urlify@@	corago:C18_Author_Role	corago:CP2_carried_out_role
EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	RuoloResponsabileEspressionePersonaggio	SWT_DOC_BRA_CAN/@@SWT_DOC_BRA_CAN.ID_DOCBRA_CAN urlify@@	corago:C6_Performer_Role	corago:CP2_carried_out_role
EventoEspressioneDocumento	DOCUMENTI/EVN/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F30_Publication_Event	EditoreDocumento	DOCUMENTI/@@DOCUMENTI.ID_DOC@@/EDITORE	crm:E39_Actor	crm:P14_carried_out_by

EventoEspressione Documento	DOCUMENTI/EVN/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F30_Publication_Event	DataEdizione	DOCUMENTI/DATA/@@DOCUMENTI.ID_DOC urlify@@	crm:E52_Time-Span	crm:P4_has_time-span
EventoEspressione Documento	DOCUMENTI/EVN/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F30_Publication_Event	EspressioneDocumento	SWT_DOC_BRA/@@SWT_DOC_BRA.ID_DOC_BRA urlify@@	frbroo:F24_Publication_Expression	crm:P94_has_created
EventoEspressione Documento	DOCUMENTI/EVN/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F30_Publication_Event	LuogoEdizione	DOCUMENTI/LUOGO/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F9_Place	crm:P7_took_place_at
EventoEspressione Documento	DOCUMENTI/EVN/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F30_Publication_Event	PaeseEdizione	DOCUMENTI/PAESE/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F9_Place	crm:P7_took_place_at
ExplicitDocumentoStruttura	DOC_STRUTTURA/DS_EXPLICIT/@@DOC_STRUTTURA.ID_DS urlify@@	crm:E33_Linguistic_Object	TipoExplicitDocumentoStruttura	DOC_STRUTTURA/TYP/EXPLICIT	crm:E55_Type	crm:P2_has_type
InventarioItem	ITEM/INV/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MINVENTARIO urlify@@	crm:E42_Identifier	TipoinventarioItem	ITEM/TYP/INV	corago:E55_Type	crm:P2_has_type
Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	frbroo:F5_Item	CollocazioneItem	ITEM/COL/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MCOLLOC urlify@@	crm:E42_Identifier	crm:P1_is_identified_by
Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	frbroo:F5_Item	Documento	DOCUMENTI/@@DOCUMENTI.ID_DOC urlify@@	frbroo:F3_Manifestation_Product_Type	frbroo:R7_is_example_of
Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	frbroo:F5_Item	Rism	RISM/@@RISM.ID_RISM@@	crm:E39_Actor	crm:P49_has_former_or_current_keeper
Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	frbroo:F5_Item	InventarioItem	ITEM/INV/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MINVENTARIO urlify@@	crm:E42_Identifier	crm:P1_is_identified_by
Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	frbroo:F5_Item	ItemColloc	ITEM/MCOLLOC/@@SWT_DOC_RISM.ID_MBIBLIO@@	crm:E62_String	crm:P3_has_note
Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	frbroo:F5_Item	ItemURL	ITEM/MBIB_URL/@@SWT_DOC_RISM.ID_MBIBLIO@@	crm:E42_Identifier	corago:CP15_has_URL
Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	frbroo:F5_Item	NotelItem	ITEM/NOTE/@@SWT_DOC_RISM.MBIB_NOTE urlify@@	crm:E62_String	crm:P3_has_note
Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	frbroo:F5_Item	OlimItem	ITEM/OLI/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MOLIM urlify@@	crm:E42_Identifier	crm:P1_is_identified_by

ItemPer mLink	ITEM/MBIB_ID/@@SWT_DO C_RISM.ID_MBIOLIO@@	crm:E42_Ide ntifier	Item	ITEM/@@SWT_DOC_RISM.ID_ MBIBLIO@@	frbroo:F5_It em	crm:P48i_is_p referred_iden tifier_of
ItemURL	ITEM/MBIB_URL/@@SWT_ DOC_RISM.ID_MBIOLIO@@	crm:E42_Ide ntifier	TipoURLItem	ITEM/TYP/URL	corago:E55_ Type	crm:P2_has_t ype
LuogoEdi zione	DOCUMENTI/LUOGO/@@D OCUMENTI.ID_DOC urlify@ @	frbroo:F9_Pl ace	LuogoEdizion ePA	LUOGO/@@DOCUMENTI.DOC_ LUOGO urlify@@	crm:E44_Pla ce_Appellati on	crm:P87_is_id entified_by
PaeseEdi zione	DOCUMENTI/PAESE/@@DO OCUMENTI.ID_DOC urlify@@	frbroo:F9_Pl ace	Nazione	TADE/PAE/@@TADE.ID_TADE urlify@@	crm:E48_Pla ce_Name	crm:P87_is_id entified_by
Produzio neDocum ento	DOCUMENTI/CPE/@@DOCU MENTI.ID_DOC urlify@@	frbroo:F32_ Carrier_Pro duction_Eve nt	DataEdizione	DOCUMENTI/DATA/@@DOCU MENTI.ID_DOC urlify@@	crm:E52_Ti me-Span	crm:P4_has_t ime-span
Produzio neDocum ento	DOCUMENTI/CPE/@@DOCU MENTI.ID_DOC urlify@@	frbroo:F32_ Carrier_Pro duction_Eve nt	LuogoEdizion e	DOCUMENTI/LUOGO/@@DOC UMENTI.ID_DOC urlify@@	frbroo:F9_Pl ace	crm:P7_took_ place_at
Produzio neDocum ento	DOCUMENTI/CPE/@@DOCU MENTI.ID_DOC urlify@@	frbroo:F32_ Carrier_Pro duction_Eve nt	Documento	DOCUMENTI/@@DOCUMENTI.I D_DOC urlify@@	frbroo:F3_ Manifestati on_Product _Type	frbroo:R26_p roduced_thin gs_of_type
Produzio neDocum ento	DOCUMENTI/CPE/@@DOCU MENTI.ID_DOC urlify@@	frbroo:F32_ Carrier_Pro duction_Eve nt	EditoreDocu mento	DOCUMENTI/@@DOCUMENTI.I D_DOC@@/EDITORE	crm:E39_Ac tor	crm:P14_carri ed_out_by
Struttura Documen to	DOC_STRUTTURA/@@DOC_ STRUTTURA.ID_DS urlify@ @	frbroo:F23_ Expression_ Fragment	DataStruttur aDocumento	DOC_STRUTTURA/DATA/@@D OC_STRUTTURA.ID_DS urlify@ @@@	crm:E52_Ti me-Span	crm:P4_has_t ime-span
Struttura Documen to	DOC_STRUTTURA/@@DOC_ STRUTTURA.ID_DS urlify@ @	frbroo:F23_ Expression_ Fragment	ExplicitDocu mentoStruttu ra	DOC_STRUTTURA/DS_EXPLICIT/ @@DOC_STRUTTURA.ID_DS ur lify@@	crm:E33_Lin guistic_Obje ct	crm:P128_car ries
Struttura Documen to	DOC_STRUTTURA/@@DOC_ STRUTTURA.ID_DS urlify@ @	frbroo:F23_ Expression_ Fragment	LuogoStruttu ra	DOC_STRUTTURA/DS_LUOGO/ @@DOC_STRUTTURA.ID_DS ur lify@@	frbroo:F9_Pl ace	crm:P7_took_ place_at
Struttura Documen to	DOC_STRUTTURA/@@DOC_ STRUTTURA.ID_DS urlify@ @	frbroo:F23_ Expression_ Fragment	StrutturaDoc umentoOrdin ale	DOC_STRUTTURA/@@DOC_STR UTTURA.ID_DS@@/@@DOC_S TRUTTURA.DS_NPRO urlify@@	crm:E60_Nu mber	corago:CPAN 1_priority
Struttura Documen to	DOC_STRUTTURA/@@DOC_ STRUTTURA.ID_DS urlify@ @	frbroo:F23_ Expression_ Fragment	StrutturaDoc umentoPagin a	DOC_STRUTTURA/PAG/@@DO C_STRUTTURA.DS_PAGINA urlif y@@	crm:E62_Str ing	crm:P3_has_ note
Struttura Documen to	DOC_STRUTTURA/@@DOC_ STRUTTURA.ID_DS urlify@ @	frbroo:F23_ Expression_ Fragment	TestoDocum entoStruttura	DOC_STRUTTURA/DS_TESTO/@ @DOC_STRUTTURA.ID_DS urlif y@@	crm:E33_Lin guistic_Obje ct	crm:P128_car ries
Struttura Documen to	DOC_STRUTTURA/@@DOC_ STRUTTURA.ID_DS urlify@ @	frbroo:F23_ Expression_ Fragment	TipoStruttura Documento	TADE/TPR&@@TADE.ID_TADE urlify@@	corago:E55_ Type	crm:P2_has_t ype

Struttura Documento	DOC_STRUTTURA/@@DOC_STRUTTURA.ID_DS urlify@@	frbroo:F23_Expression_Fragment	RuoloResponsabileStruttura	SWT_DOC_STRUT_RESP/@@SWT_DOC_STRUT_RESP.ID_DSR urlify@@	corago:C2_Actor_Role	corago:CP2_carried_out_role
TestoDocumentoStruttura	DOC_STRUTTURA/DS_TESTO/@@DOC_STRUTTURA.ID_DS urlify@@	crm:E33_Linguistic_Object	TipoDocumentoStruttura	DOC_STRUTTURA/TYP/TESTO	crm:E55_Type	crm:P2_has_type
RuoloResponsabileDocumento	SWT_DOC_RESP/@@SWT_DOC_RESP.ID_DOC_RESP@@	corago:C2_Actor_Role	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CP3_carried_out_actor
RuoloResponsabileStruttura	SWT_DOC_STRUT_RESP/@@SWT_DOC_STRUT_RESP.ID_DSR urlify@@	corago:C2_Actor_Role	RuoloResponsabileStrutturaNota	SWT_DOC_STRUT_RESP/NOME/@@SWT_DOC_STRUT_RESP.ID_DSR_NOME urlify@@	crm:E62_String	crm:P3_has_note
RuoloResponsabileStruttura	SWT_DOC_STRUT_RESP/@@SWT_DOC_STRUT_RESP.ID_DSR urlify@@	corago:C2_Actor_Role	Responsabile	RESPONS/@@RESPONS.COD_RESP urlify@@	crm:E39_Actor	corago:CP3_carried_out_actor

A.9.Mapping degli esemplari

In *Corago* l'elenco delle biblioteche in cui vengono conservati gli esemplari di Documenti (*items* in FRBR) è censito nella tabella RISM. Le rispettive collocazioni sono riportate nella tabella SWT_DOC_RISM e rappresentata attraverso la classe frbroo:F5_Item come illustrato nel diagramma A.9.

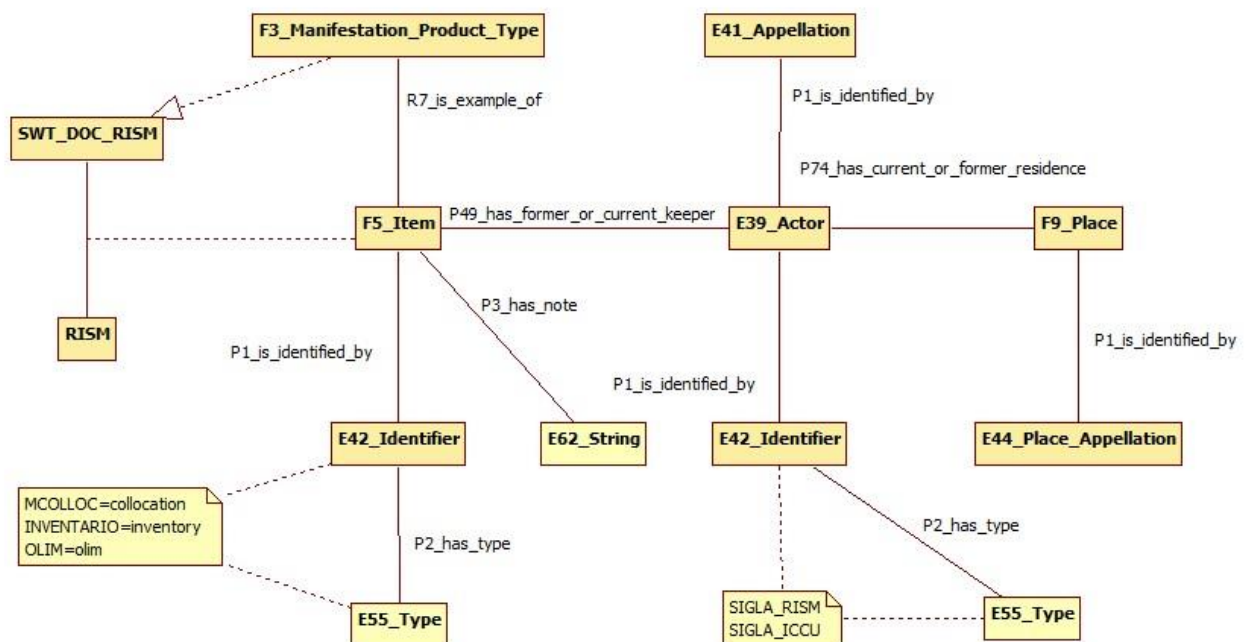


Diagramma A.9: Mapping esemplari (*items*)

A.9.1. Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
Item	frbroo:F5_Item	ITEM/@@SWT_DOC_RISM.ID_MBIBLIO@@	rdfs:label	DOCUMENTI.DOC_SIGLA - RISM.SIGLA_RISM - SWT_DOC_RISM.MCOLLOC
Rism	crm:E39_Actor	RISM/@@RISM.ID_RISM@@	rdfs:label	RISM.SIGLA_RISM
SiglaRism	crm:E42_Identifier	RISM/SIGLA/@@RISM.SIGLA_RISM urlify@@	rdfs:label	RISM.SIGLA_RISM
RismISTURL	crm:E42_Identifier	RISM/RISMURL/@@RISM.ID_RISM@@	rdfs:label	RISM.URL_IST
LuogoRism	frbroo:F9_Place	RISM/LUOGO/@@RISM.CITTA urlify@@	rdfs:label	RISM.CITTA
CollocazioneItem	crm:E42_Identifier	ITEM/COL/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MCOLLOC urlify@@	rdfs:label	SWT_DOC_RISM.MCOLLOC
Inventarioltem	crm:E42_Identifier	ITEM/INV/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MINVENTARIO urlify@@	rdfs:label	SWT_DOC_RISM.MINVENTARIO
ItemPermLink	crm:E42_Identifier	ITEM/MBIB_ID/@@SWT_DOC_RISM.ID_MBIBLIO@@	rdfs:label	SWT_DOC_RISM.MBIB_ID
ItemURL	crm:E42_Identifier	ITEM/MBIB_URL/@@SWT_DOC_RISM.ID_MBIBLIO@@	rdfs:label	SWT_DOC_RISM.MBIB_URL
ItemURL	crm:E42_Identifier	ITEM/MBIB_URL/@@SWT_DOC_RISM.ID_MBIBLIO@@	rdfs:label	RISM.RURL1 & SWT_DOC_RISM.MBIB_URL & RISM.RURL2
OlimItem	crm:E42_Identifier	ITEM/OLI/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MOLIM urlify@@	rdfs:label	SWT_DOC_RISM.MOLIM
ItemColloc	crm:E62_String	ITEM/MCOLLOC/@@SWT_DOC_RISM.ID_MBIBLIO@@	rdfs:label	SWT_DOC_RISM.MCOLLOC

A.9.2. Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
CollocazioneItem	ITEM/COL/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MCOLLOC urlify@@	crm:E42_Identifier	TipoCollocazioneItem	ITEM/TYP/COL	corago:E55_Type	crm:P2_has_type
Inventarioltem	ITEM/INV/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MINVENTARIO urlify@@	crm:E42_Identifier	TipoInventarioltem	ITEM/TYP/INV	corago:E55_Type	crm:P2_has_type

Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo: F5_It em	Collocaz ionelte m	ITEM/COL/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.M COLLOC urlify@@	crm:E42_Identi fier	crm:P1_is_identi fied_by
Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo: F5_It em	Docum ento	DOCUMENTI/@@DOCUMENTI.ID_ DOC urlify@@	frbroo:F3_Man ifestation_Pro duct_Type	frbroo:R7_is_ex ample_of
Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo: F5_It em	Rism	RISM/@@RISM.ID_RISM@@	crm:E39_Actor	crm:P49_has_fo rmer_or_current _keeper
Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo: F5_It em	Inventa rioltem	ITEM/INV/@@RISM.SIGLA_RISM urlify@@/@@SWT_DOC_RISM.MI NVENTARIO urlify@@	crm:E42_Identi fier	crm:P1_is_identi fied_by
Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo: F5_It em	ItemCol loc	ITEM/MCOLLOC/@@SWT_DOC_RI SM.ID_MBI BLIO@@	crm:E62_String	crm:P3_has_not e
Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo: F5_It em	ItemUR L	ITEM/MBIB_URL/@@SWT_DOC_R ISM.ID_MBI BLIO@@	crm:E42_Identi fier	corago:CP15_ha s_URL
Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo: F5_It em	Notelte m	ITEM/NOTE/@@SWT_DOC_RISM. MBIB_NOTE urlify@@	crm:E62_String	crm:P3_has_not e
Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo: F5_It em	Olimlte m	ITEM/OLI/@@RISM.SIGLA_RISM u rlify@@/@@SWT_DOC_RISM.MO LIM urlify@@	crm:E42_Identi fier	crm:P1_is_identi fied_by
ItemP ermLi nk	ITEM/MBIB_ID/@@SWT_DOC_RIS M.ID_MBI BLIO@@	crm:E4 2_Iden tifier	Item	ITEM/@@SWT_DOC_RISM.ID_MBI BLIO@@	frbroo:F5_Item	crm:P48i_is_pre ferred_identifier _of
ItemU RL	ITEM/MBIB_URL/@@SWT_DOC_R ISM.ID_MBI BLIO@@	crm:E4 2_Iden tifier	TipoURL Item	ITEM/TYP/URL	corago:E55_Ty pe	crm:P2_has_typ e
Olimlt em	ITEM/OLI/@@RISM.SIGLA_RISM u rlify@@/@@SWT_DOC_RISM.MO LIM urlify@@	crm:E4 2_Iden tifier	TipoOli mltem	ITEM/TYP/OLI	corago:E55_Ty pe	crm:P2_has_typ e
Rism	RISM/@@RISM.ID_RISM@@	crm:E3 9_Act or	IccuRis m	RISM/ICCU/@@RISM.SIGLA_ICCU urlify@@	crm:E42_Identi fier	crm:P1_is_identi fied_by
Rism	RISM/@@RISM.ID_RISM@@	crm:E3 9_Act or	LuogoRi sm	RISM/LUOGO/@@RISM.CITTA urli fy@@	frbroo:F9_Plac e	crm:P74_has_cu rrent_or_former _residence
Rism	RISM/@@RISM.ID_RISM@@	crm:E3 9_Act or	RismAp pellatio n	RISM/APP/@@RISM.DENOMINAZI ONE urlify@@	crm:E82_Actor _Appellation	crm:P1_is_identi fied_by
Rism	RISM/@@RISM.ID_RISM@@	crm:E3 9_Act or	RismIST URL	RISM/RISMURL/@@RISM.ID_RISM @@	crm:E42_Identi fier	crm:P1_is_identi fied_by

A.10.1.Mapping degli attributi

Entità Corago	Classe destinazione	Campo origine entità	Proprietà	Campo origine proprietà
Bibliografia	crm:E31_Document	BIBLIOGRAFIA/@@BIBLIOGRAFIA.ID_BIB urlify@@	rdfs:label	BIBLIOGRAFIA.SIGLA_BIB
TipoBibliografia	corago:E55_Type	BIBLIOGRAFIA/TYP&@@BIBLIOGRAFIA.TIPO_BIB urlify@@	rdfs:label	BIBLIOGRAFIA.TIPO_BIB
TitoloBibliografia	crm:E35_Title	BIBLIOGRAFIA/TIT_ORD/@@BIBLIOGRAFIA.ID_BIB urlify@@	rdfs:label	BIBLIOGRAFIA.TIT_ORD
BiblioURL	crm:E42_Identifier	BIBLIOGRAFIA/URL/@@BIBLIOGRAFIA.ID_BIB urlify@@	rdfs:label	BIBLIOGRAFIA.BIB_URL
DataPubblicazioneBiblio	crm:E52_Time-Span	BIBLIOGRAFIA/ANNO/@@BIBLIOGRAFIA.ID_BIB urlify@@	rdfs:label	BIBLIOGRAFIA.ANNO
PubblicazioneBiblioNote	crm:E62_String	BIBLIOGRAFIA/PAG/@@BIBLIOGRAFIA.NOTE_BIB urlify@@	rdfs:label	BIBLIOGRAFIA.NOTE_BIB
PubblicazioneBiblioPag	crm:E62_String	BIBLIOGRAFIA/PAG/@@BIBLIOGRAFIA.PAGINAZIONE urlify@@	rdfs:label	BIBLIOGRAFIA.PAGINAZIONE
LuogoPubblicazioneBiblio	frbroo:F9_Place	BIBLIOGRAFIA/LUOGO/@@BIBLIOGRAFIA.ID_BIB urlify@@	rdfs:label	BIBLIOGRAFIA.LUOGO_ED

A.10.2.Mapping delle entità associate

Entità Corago	Origine entità	Classe entità	Entità associata	Origine entità associata	Classe entità associata	Relazione
Bibliografia	BIBLIOGRAFIA/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E31_Document	BiblioURL	BIBLIOGRAFIA/URL/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E42_Identifier	corago:CP15_has_URL
Bibliografia	BIBLIOGRAFIA/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E31_Document	TipoBibliografia	BIBLIOGRAFIA/TYP&@@BIBLIOGRAFIA.TIPO_BIB urlify@@	corago:E55_Type	crm:P2_has_type
Bibliografia	BIBLIOGRAFIA/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E31_Document	TitoloBibliografia	BIBLIOGRAFIA/TIT_ORD/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E35_Title	crm:P102_has_title
Bibliografia	BIBLIOGRAFIA/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E31_Document	TitoloBibliografia	BIBLIOGRAFIA/TIT_ORD/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E35_Title	corago:CPBC20_title
Bibliografia	BIBLIOGRAFIA/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E31_Document	TitoloLibBibliografia	BIBLIOGRAFIA/TIT_LIB/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E35_Title	corago:CP16_book_title
Bibliografia	BIBLIOGRAFIA/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E31_Document	TitoloSagBibliografia	BIBLIOGRAFIA/TIT_SAG/@@BIBLIOGRAFIA.ID_BIB urlify@@	crm:E35_Title	corago:CP17_most_specific_title

Bibliograf ia	BIBLIOGRAFIA/@@BIBLI OGRAFIA.ID_BIB urlify@ @	crm:E31_Docu ment	TitoloSagBibl iografia	BIBLIOGRAFIA/TIT_SAG/@@ BIBLIOGRAFIA.ID_BIB urlif y@@	crm:E35_Title	corago:CPBC2 0_title
Bibliograf ia	BIBLIOGRAFIA/@@BIBLI OGRAFIA.ID_BIB urlify@ @	crm:E31_Docu ment	RinvioBiblioB rano	LEGAMI/RIN_BRA/@@SWT _BRANI_BIB.ID_RINV urlify @@	corago:C19_Bibli ographic_Refere nce	corago:CP5_r efers_with
Bibliograf ia	BIBLIOGRAFIA/@@BIBLI OGRAFIA.ID_BIB urlify@ @	crm:E31_Docu ment	RinvioBiblio	LEGAMI/RIN/@@SWT_DOC _BIB.ID_RINV urlify@@	corago:C19_Bibli ographic_Refere nce	corago:CP5_r efers_with
Bibliograf ia	BIBLIOGRAFIA/@@BIBLI OGRAFIA.ID_BIB urlify@ @	crm:E31_Docu ment	RinvioBiblioS pettacolo	LEGAMI/RIN_SPE/@@RIN BIB.ID_RINV urlify@@	corago:C5_Link	corago:CP5_r efers_with
Pubblicaz ioneBibli o	BIBLIOGRAFIA/PUBB/@@ BIBLIOGRAFIA.ID_BIB urlify@@	frbroo:F30_Pub lication_Event	DataPubblica zioneBiblio	BIBLIOGRAFIA/ANNO/@@B IBLIOGRAFIA.ID_BIB urlify@ @	crm:E52_Time- Span	crm:P4_has_t ime-span
Pubblicaz ioneBibli o	BIBLIOGRAFIA/PUBB/@@ BIBLIOGRAFIA.ID_BIB urlify@@	frbroo:F30_Pub lication_Event	LuogoPublic azioneBiblio	BIBLIOGRAFIA/LUOGO/@@ BIBLIOGRAFIA.ID_BIB urlify @@	frbroo:F9_Place	crm:P7_took_ place_at
Pubblicaz ioneBibli o	BIBLIOGRAFIA/PUBB/@@ BIBLIOGRAFIA.ID_BIB urlify@@	frbroo:F30_Pub lication_Event	Pubblicazion eBiblioNote	BIBLIOGRAFIA/PAG/@@BIB LIOGRAFIA.NOTE_BIB urlify @@	crm:E62_String	crm:P3_has_ note
Pubblicaz ioneBibli o	BIBLIOGRAFIA/PUBB/@@ BIBLIOGRAFIA.ID_BIB urlify@@	frbroo:F30_Pub lication_Event	Pubblicazion eBiblioPag	BIBLIOGRAFIA/PAG/@@BIB LIOGRAFIA.PAGINAZIONE u rlify@@	crm:E62_String	crm:P3_has_ note
RinvioBib lio	LEGAMI/RIN/@@SWT_D OC_BIB.ID_RINV urlify@ @	corago:C19_Bib liographic_Refere nce	RinvioBiblioI D	LEGAMI/RIN/ID/@@SWT_D OC_BIB.ID_RINV urlify@@	crm:E42_Identifi er	crm:P1_is_ide ntified_by
RinvioBib lio	LEGAMI/RIN/@@SWT_D OC_BIB.ID_RINV urlify@ @	corago:C19_Bib liographic_Refere nce	RinvioBiblioN ote	LEGAMI/RIN/NOTE/@@SW T_DOC_BIB.ID_RINV urlify @@	crm:E62_String	crm:P3_has_ note
RinvioBib lio	LEGAMI/RIN/@@SWT_D OC_BIB.ID_RINV urlify@ @	corago:C19_Bib liographic_Refere nce	RinvioBiblioO rdinale	LEGAMI/RIN/ORD/@@SWT _DOC_BIB.ID_RINV urlify@ @	crm:E60_Numbe r	corago:CPAN 1_priority
RinvioBib lio	LEGAMI/RIN/@@SWT_D OC_BIB.ID_RINV urlify@ @	corago:C19_Bib liographic_Refere nce	RinvioBiblioU RL	LEGAMI/RIN/ID/URL/@@S WT_DOC_BIB.ID_RINV urlif y@@	crm:E42_Identifi er	corago:CP15_ has_URL
RinvioBib lio	LEGAMI/RIN/@@SWT_D OC_BIB.ID_RINV urlify@ @	corago:C19_Bib liographic_Refere nce	Documento	DOCUMENTI/@@DOCUME NTI.ID_DOC urlify@@	frbroo:F3_Manif estation_Produc t_Type	corago:CP6_r efers_to
RinvioBib lioBrano	LEGAMI/RIN_BRA/@@S WT_BRANI_BIB.ID_RINV urlify@@	corago:C19_Bib liographic_Refere nce	RinvioBiblioB ranoNota	LEGAMI/RIN_BRA/NOTE/@@ SWT_BRANI_BIB.ID_RINV urlify@@	crm:E62_String	crm:P3_has_ note
RinvioBib lioBrano	LEGAMI/RIN_BRA/@@S WT_BRANI_BIB.ID_RINV urlify@@	corago:C19_Bib liographic_Refere nce	RinvioBiblioB ranoOrdinale	LEGAMI/RIN_BRA/ORD/@@ SWT_BRANI_BIB.ID_RINV urlify@@	crm:E60_Numbe r	corago:CPAN 1_priority

RinvioBiblioSpettacolo	LEGAMI/RIN_SPE/@@RIN_BIB.ID_RINV urlify@	corago:C5_Link	RinvioBiblioSpettacoloNota	LEGAMI/RIN_SPE/NOTE/@@RIN_BIB.ID_RINV urlify@	crm:E62_String	crm:P3_has_note
RinvioBiblioSpettacolo	LEGAMI/RIN_SPE/@@RIN_BIB.ID_RINV urlify@	corago:C5_Link	RinvioBiblioSpettacoloOrdinale	LEGAMI/RIN_SPE/ORD/@@RIN_BIB.ID_RINV urlify@	crm:E60_Number	corago:CPAN1_priority
RinvioBiblioSpettacolo	LEGAMI/RIN_SPE/@@RIN_BIB.ID_RINV urlify@	corago:C5_Link	Spettacolo	SPETTACOLI/@@SPETTACOLI.ID_SPET urlify@	frbroo:F31_Performance	corago:CP6_refers_to

A.11. Modello relazionale Corago

In conclusione, si riporta nel diagramma A.11 una rappresentazione sinottica del modello logico della base dati relazionale di *Corago*. Vengono riportate le principali entità del dominio: Brano, Spettacolo, Documento e Responsabile assieme alle rispettive relazioni. L'insieme di queste entità costituiscono il perimetro informativo interessato dal processo di trasposizione descritto nei paragrafi precedenti.

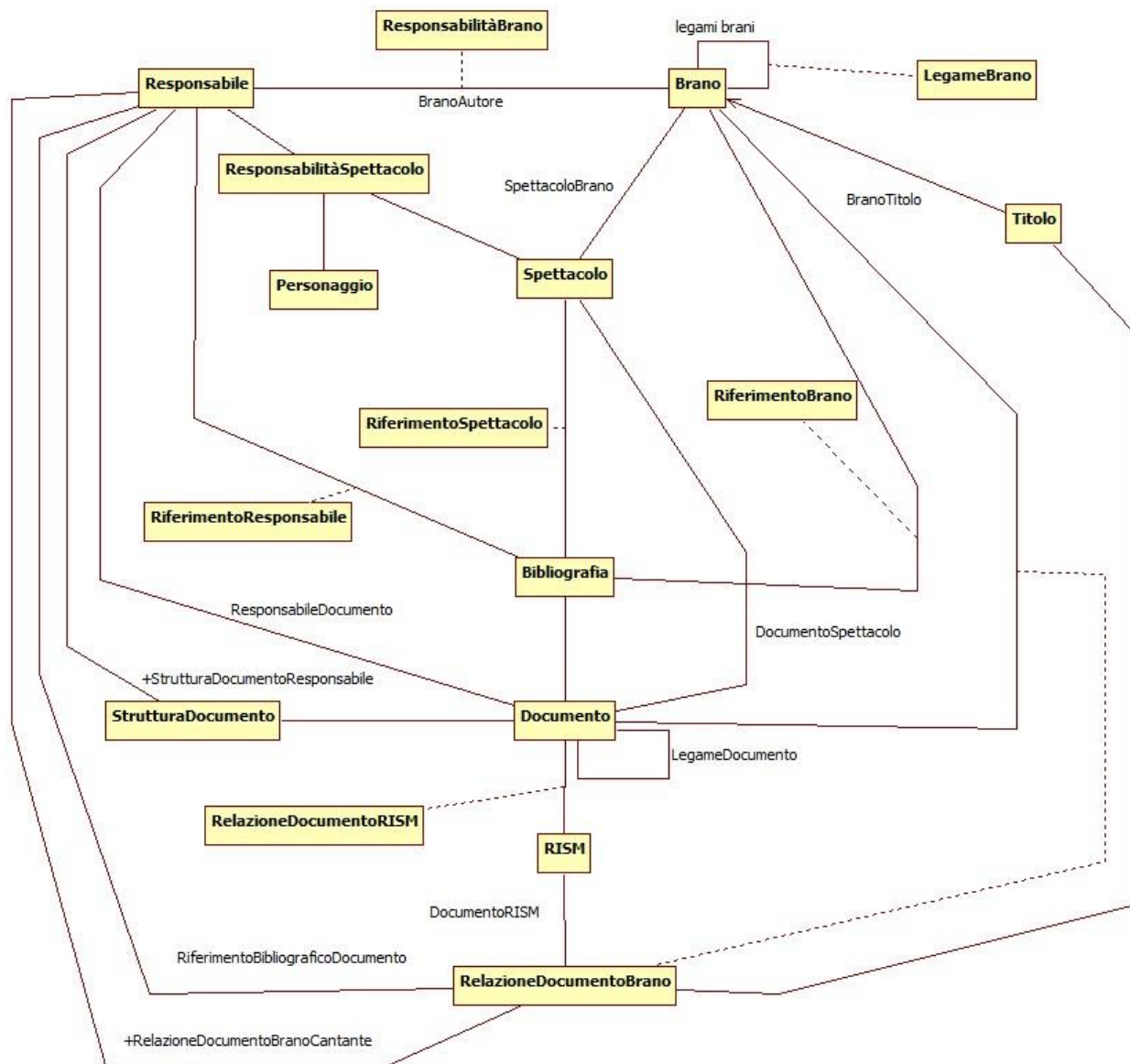


Diagramma A.11: Modello concettuale Corago

B. Architettura del sistema sperimentale

B.1. Architettura del sistema

L'architettura del sistema che supporta il processo di conversione dei contenuti dell'archivio dal modello relazionale al modello semantico e la loro consultazione è illustrata nel diagramma B.1.

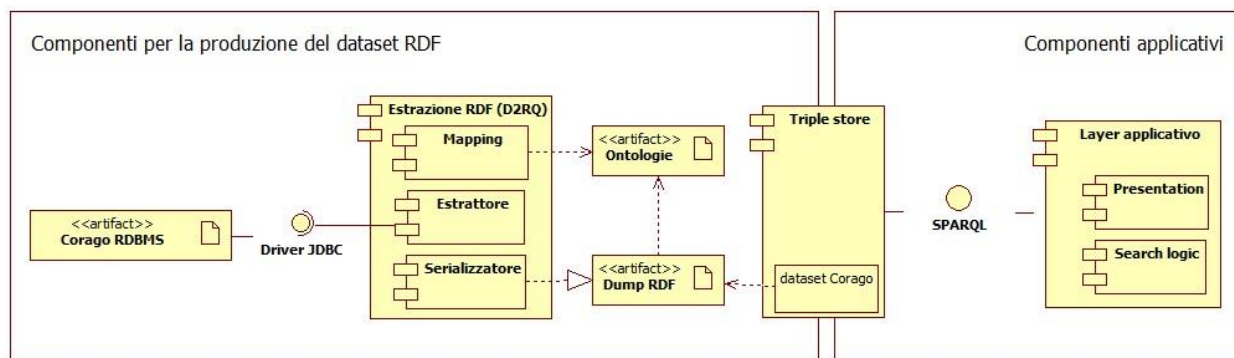


Diagramma B.1: Architettura di sistema

Il sistema è articolato in due sezioni: l'insieme dei componenti impiegati per l'estrazione e trasformazione dei contenuti in formato RDF e l'applicazione per la consultazione. Entrambe utilizzano componenti software basati sul linguaggio Java mentre la definizione del modello concettuale ed i relativi file di configurazione utilizzano la famiglia dei linguaggi definiti dal W3C per il web semantico: da RDF a OWL.

La fonte è costituita dal RDMS di *Corago*: un file Microsoft Access formato '97 che viene letto attraverso una connessione di tipo JDBC. I componenti per la generazione del file di *dump* RDF è interamente basata sulla suite D2RQ.

Il *repository* delle triple RDF (*triple store*) è basato sul prodotto Blazegraph nella sua ultima versione 2.1.5 che viene installata come applicazione di tipo WAR all'interno di un server Apache Tomcat versione 8. Il *repository* funge da sorgente dati per l'applicazione di consultazione che accede ai contenuti attraverso il relativo *endpoint* SPARQL 1.1.

L'applicazione di consultazione è stata sviluppata secondo il *pattern* Model View Controller (MVC) partendo dalla soluzione open source LodView¹⁸⁸.

¹⁸⁸ La soluzione è pubblicata sul *repository* GitHub all'indirizzo: <https://github.com/dvcama/LodView> (verificato 29/09/2019).

B.2. Applicazione di consultazione

Le componenti dell'applicazione di consultazione dei contenuti sono rappresentate nel diagramma B.2.

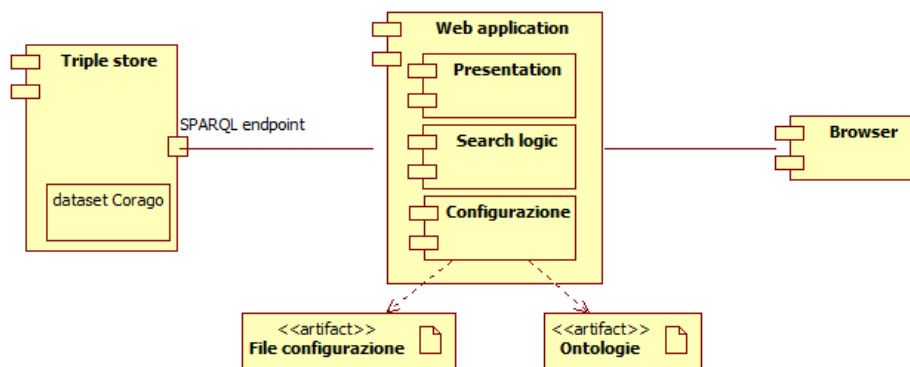


Diagramma B.2: Architettura applicazione di consultazione

L'applicazione si compone di tre componenti principali: due di natura funzionale che implementano le logiche di ricerca e di presentazione ed una di natura infrastrutturale che gestisce le configurazioni relative al modello concettuale di riferimento e applicative. Il *pattern* software MVC prevede la separazione tra le componenti che definiscono il modello delle entità trattate dalle logiche applicative delegate ai *controller* e le viste utente (*views*). L'applicazione sviluppata eredita dal progetto LodView l'uso del *framework* Spring MVC¹⁸⁹.

Nella appendice C vengono illustrati alcuni dei componenti software più significativi introdotti per supportare le funzionalità di consultazione e ricerca introdotte durante la sperimentazione.

L'applicazione di consultazione è utilizzabile attraverso un comune browser web ed è interamente *web based* dialogando sia col *backend* che coi client attraverso il protocollo HTTP. L'applicazione è ospitata su un server separato rispetto al *triple store*, anche questo di tipo Apache Tomcat 8.

La scelta di separare fisicamente i due strati: *backend* dati e *frontend* applicativo, deriva dalla necessità di isolare il secondo dal carico computazionale generato dal primo in fase di manutenzione delle inferenze (*closure*) della KB e die esecuzione degli algoritmi di estrazione dell'informazione descritti nel capitolo 4.

B.3. Indicizzazione full-text

Il componente dedicato alla costruzione dell'indice testuale secondo le modalità descritte nel paragrafo 2.6 è rappresentato nelle sue componenti fondamentali nel diagramma A.3.

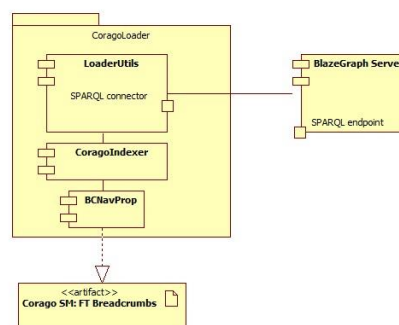


Diagramma B.3: Componente per l'indicizzazione full text

Il processo di creazione dell'indice avviene a valle del caricamento dei contenuti all'interno del *triple store*. L'indice è costruito in base alle proprietà che specificano quali contenuti del grafo devono entrarne a far parte (proprietà di tipo FTBreadcrumbs). Il componente software dedicato prevede la lettura di queste proprietà dal modello concettuale, l'estrazione i contenuti testuali dal *dataset* per tutte le risorse da indicizzare e il riversamento delle triple di indice create nuovamente nel *repository* RDF (vedi paragrafo 2.6.2.3). La logica di costruzione delle triple dell'indice è implementata nel modulo CoragoLoader, mentre il modulo LoaderUtils è dedicato alla negoziazione con il *repository*. L'intero componente è progettato in modo da poter essere eseguito come processo autonomo (*standalone*) che come modulo integrabile nei processi di indicizzazione del *repository* stesso. In questo modo l'indicizzazione può avvenire contestualmente al caricamento dei contenuti.

B.4. Annotazione ed analisi dei contenuti testuali

Le sperimentazioni di estrazione automatica dei contenuti descritti nel paragrafo 4.2 sono state condotte attraverso lo sviluppo del componente Corago NLP la cui architettura è illustrata nel diagramma A.4.

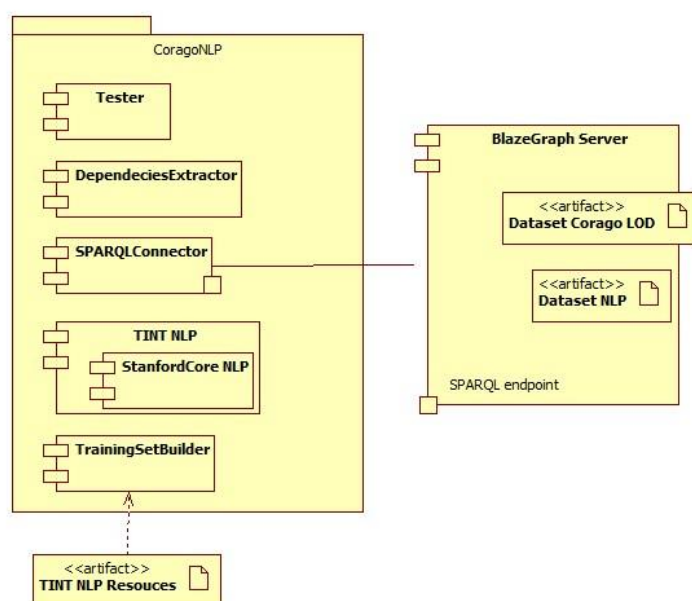


Diagramma A.4: Componente per l'annotazione NLP

Il componente utilizza per tutte le funzionalità di annotazione lo *stack* software Stanford CoreNLP nella versione localizzata per l'italiano dalla Fondazione Bruno Kessler col progetto TINT. Per l'approntamento dei training set necessari alla sperimentazione è stato realizzato un componente dedicato il TrainingSetBuilder con il compito di estrarre i contenuti dalle fonti e preparare l'output nei formati previsti dagli *stack* NLP.

A partire dalle funzionalità esposte dallo Stanford CoreNLP è stata implementata la *business logic* necessaria alla sperimentazione nei moduli: DependenciesExtractor e Tester. La negoziazione dei contenuti coi due *dataset* interessati dalla sperimentazione è delegata al modulo SPARQLConnector.

Il componente è stato disegnato per operare esclusivamente come un processo autonomo.

C. Componenti software

L'appendice presenta alcune sezioni di codice relative alle caratteristiche del sistema sperimentale più significative presentate nella tesi. Vengono riportate come esempi di una concreta implementazione delle soluzioni proposte nei diversi ambiti affrontati durante la sperimentazione. L'obiettivo è illustrare come queste possano essere impiegate all'interno di una architettura di tipo J2EE in prospettiva di una loro applicabilità in altri contesti.

Il codice riportato è in linguaggio Java, l'editor utilizzato è Eclipse (versione Mars.2).

C.1. Indicizzazione full-text

L'implementazione della procedura necessaria alla costruzione delle triple che costituiscono l'indice full-text descritto nel paragrafo 2.6 è contenuta in un progetto specifico denominato Corago Loader.

La classe principale CoragoLoader provvede all'estrazione, assemblaggio e caricamento (da cui il nome) delle triple dell'indice nel *repository* target.

La classe è articolata in tre metodi principali:

- `createFullTextIndex(String)`: metodo per la generazione del file completo dell'indice;
- `getIndexingProperties(String)`: metodo per l'estrazione dei contenuti da indicizzare;
- `loadFTIndexFile(String)`: metodo, con *overloads*, per il caricamento del file.

Il primo contiene la procedura principale, gli altri due le funzioni ausiliarie.

C.1.1. Costruzione dell'indice: il metodo '`createFullTextIndex(String)`'

Il metodo implementa una procedura strutturata in due fasi: la preparazione dei dati e la composizione delle triple dell'indice.

La prima consiste nelle operazioni preliminari necessarie a predisporre un *dataset* temporaneo che viene utilizzato come guida per la successiva indicizzazione. Il listato C.1.1 riporta il codice relativo a questo passaggio iniziale.


```
//STEP 1.3 - generazione del solutionset usato come indice guida
if(step<2)
{
    System.out.println("***SuperIndiceFT: inizio generazione solutionSet");
    String solutionsetIndex = "INSERT INTO %solutionSet0"+
        " SELECT distinct ?s WHERE{?s a ?class.FILTER (?class IN (<http://erlangen-
crm.org/efrbroo/F3_Manifestation_Product_Type>,<http://erlangen-crm.org/efrbroo/F31_Performance>,<http://erlangen-
crm.org/efrbroo/F15_Complex_Work>,<http://erlangen-crm.org/efrbroo/F14_Individual_Work>))}";
    try {
        RemoteRepository namedRep = repo.getRepositoryForNamespace(namespace);
        IPreparedSparqlUpdate updQ = namedRep.prepareUpdate(solutionsetIndex);
        updQ.evaluate();

        System.out.println("***SuperIndiceFT: generato solutionSet0");
    } catch (Exception e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }
}
```

Listato C.1.1

Occorre sottolineare come sia questa la fase in cui vengono selezionate, in base alla classe di appartenenza, le risorse che verranno indicizzate. Come requisito per il sistema sperimentale si è deciso di indicizzare solo le entità fondamentali del dominio del teatro d'opera: frbroo:F3_Manifestation_Product_Type (ovvero i libretti), frbroo:F31_Performance (gli spettacoli), frbroo:F15_Complex_Work e frbroo:F14_Individual_Work (ovvero le opere). Questo significa che l'indice contiene solo occorrenze relative queste tipologie di entità. Di conseguenza le ricerche che lo utilizzano produrranno risultati coerenti con questa impostazione. Si noti che qualora si vogliano includere altre tipologie di risorse sarà sufficiente inserirle in questo insieme iniziale di classi per ottenerne l'indicizzazione.

La fase di costruzione delle triple è articolata in più passaggi: l'inizializzazione dell'output (sia esso su file o con salvataggio diretto nel *repository*), il caricamento del *dataset* delle risorse da indicizzare (creato nella fase precedente) e la costruzione delle singole triple. Il listato C.1.2 riporta il codice relativo a quest'ultimo passaggio. Si noti come l'individuazione dei contenuti per ciascuna risorsa venga ottenuta attraverso una interrogazione SPARQL che individua tutte le risorse associate ad essa mediante la proprietà corago:CPA4_ft_breadcrumb.

```

//iterazione sul set per ogni risorsa da indicizzare
while(tlIt.hasNext())
{
    String currURI = tlIt.next();
    currTarget = "<" + currURI + ">";
    procCnt++;
    //compongo la stringa di entità source dal %solutionSet
    String sources = "";
    searchQuery = "prefix bds: <http://www.bigdata.com/rdf/search#> prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> prefix ecrm: <http://erlangen-crm.org/current/> prefix frbroo: <http://erlangen-crm.org/efrbroo/> prefix corago: <http://corago.unibo.it/sm/> prefix xsd: <http://www.w3.org/2001/XMLSchema#> prefix geosparql: <http://www.opengis.net/ont/geosparql#> prefix geovocab: <http://geovocab.org/geometry#> prefix fn: <http://www.w3.org/2005/xpath-functions#> "
    + "SELECT distinct (?s0 as ?s) (<http://corago.unibo.it/sm/CPA5_ft_indexed_by> as ?p) (concat(str(?e0),\"&&\",str(?l0)) as ?obj) WHERE"
    + "      { { ?e0 rdfs:label ?l0. ?s0 <http://corago.unibo.it/sm/CPA4_ft_breadcrumb> ?e0.} UNION { ?e0 ecrm:P131_is_identified_by/^ecrm:P131_is_identified_by/rdfs:label ?l0. ?s0 <http://corago.unibo.it/sm/CPA4_ft_breadcrumb> ?e0. }"
    + "filter(sameTerm(?s0,\"+currTarget+\")) } ";
    try {
        res = repo.getRepositoryForNamespace(namespace).prepareTupleQuery(searchQuery).evaluate();
        while(res.hasNext())
        {
            BindingSet aB = res.next();
            sources += aB.getValue((String) "obj").stringValue().replace("\\",
            "/").replaceAll("\\r\\n|\\r|\\n", " ");
            if(res.hasNext()) sources += "###";
        }
        //inserisco le proprietà per l'ordinamento in coda all'indice
        Map<String, String> indexingProp = getIndexingProperties(currURI);
        String indexedProp = "";
        try{
            for(Entry<String, String> aE: indexingProp.entrySet())
            {
                String aVal = aE.getValue().replace("\\", "").replace("\n", "").replace("\r", "");
                aVal = aVal.replaceAll("\\r\\n|\\r|\\n", " ");
                indexedProp+=aE.getKey()+"&&"+aVal+"###";
            }
        }
        catch (Exception e2) {e2.printStackTrace(); }

        String singleRow = currTarget + " <http://corago.unibo.it/sm/CPA5_ft_indexed_by> \"\" +sources.replace("\\",
        ""))+\"#&&"+indexedProp+"\".";

        //aggiungo al file
        writer.write(singleRow+String.format("%n"));
        //se richiesto inserisco direttamente nel repository
        if(loadFTData)
            loadFTIndexTriple(singleRow);

        //3) flush del batch per singolo target sul file di output
        writer.flush();
        Integer prog = (procCnt*100)/batchSize;
        System.out.println("***SuperIndiceFT: "+ prog.toString() + "% scrittura risorsa:"+
currTarget);

```

Listato C.1.2

```

private static Map<String, String> getIndexingProperties(String entityURI)
{
    Map<String, String> ret = new HashMap<String, String>();
    String searchIRI = entityURI;
    List<String> queryList = new ArrayList<String>();

    String searchQuery = "PREFIX corago: <http://corago.unibo.it/sm/> PREFIX sesame:
<http://www.openrdf.org/schema/sesame#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> prefix rdfs:
<http://www.w3.org/2000/01/rdf-schema#> prefix ecrm: <http://erlangen-crm.org/current/> "
        + "SELECT ?s ?p ?o {"
        + "{SELECT distinct (<" + searchIRI + "> as ?s) (rdfs:label as ?p) ?o {<" + searchIRI + "> rdfs:label ?o.
FILTER (!isBlank(?o)).}}"
        + "UNION { SELECT distinct(<" + searchIRI + "> as ?s) (rdf:type as ?p) ?o {"
        + "<" + searchIRI + "> a ?o. FILTER (!isBlank(?o)).} UNION { <><><>}"
        + "}"
        + "UNION { SELECT distinct (<" + searchIRI + "> as ?s) (ecrm:P4_has_time-span as ?p) ?o {<" +
searchIRI + "> <http://erlangen-crm.org/current/P4_has_time-span>/rdfs:label|^<http://erlangen-
crm.org/efrbroo/R26_produced_things_of_type>/<http://erlangen-crm.org/current/P4_has_time-
span>/rdfs:label|corago:CP9_first_performed_in/<http://erlangen-crm.org/current/P4_has_time-span>/rdfs:label|<http://erlangen-
crm.org/efrbroo/R16i_was_initiated_by>/<http://erlangen-crm.org/current/P4_has_time-span>/rdfs:label ?o. FILTER (!isBlank(?o)).}}"
        + "UNION { SELECT distinct (<" + searchIRI + "> as ?s) (corago:indexingAuthor as ?p) ?o {<" + searchIRI
+ "> <http://erlangen-crm.org/efrbroo/R16i_was_initiated_by>*/corago:CP2_carried_out_role|<http://erlangen-
crm.org/efrbroo/R25_performed>/^<http://erlangen-crm.org/efrbroo/R17_created>/corago:CP2_carried_out_role ?role. ?role
corago:CP3_carried_out_actor/rdfs:label ?o. ?role <http://corago.unibo.it/sm/CPAN1_priority>/rdfs:label ?pri. FILTER
(regex(STR(?pri),\"1\\\", \"i\\\")))}"
        + "UNION { SELECT distinct (<" + searchIRI + "> as ?s) (corago:indexingAuthor as ?p) ?o {<" + searchIRI
+ "> <http://erlangen-crm.org/efrbroo/R16i_was_initiated_by>*/corago:CP2_carried_out_role|<http://erlangen-
crm.org/efrbroo/R25_performed>/^<http://erlangen-crm.org/efrbroo/R17_created>/corago:CP2_carried_out_role ?role. ?role
corago:CP3_carried_out_actor/rdfs:label ?o. FILTER NOT EXISTS {?role <http://corago.unibo.it/sm/CPAN1_priority>/rdfs:label ?pri.} }}"
        + "UNION { SELECT distinct (<" + searchIRI + "> as ?s) (corago:approxDate as ?p) ?o {<" + searchIRI
+ "> <http://erlangen-crm.org/efrbroo/R16i_was_initiated_by>/<http://corago.unibo.it/sm/CP13_first_performance_date>/<http://erlangen-
crm.org/current/P78_is_identified_by> ?t. ?t a <http://corago.unibo.it/sm/C16_Approximate_Time_Appellation>. ?t rdfs:label ?o }}"
        + "}"
        + "}"

    TupleQueryResult res = null;

    try {
        res = repo.getRepositoryForNamespace(namespace).prepareTupleQuery(searchQuery).evaluate();
        while(res.hasNext())
        {
            BindingSet aB = res.next();
            String property = "";
            property = aB.getValue((String) "p").stringValue();
            String val = "";
            val = aB.getValue((String) "o").stringValue().replace(".", "", "");

            if (property.toString().contains("directType")) {
                val = val;
                if (val == null)
                    val = "rdf:type";
            }
            else
                val = val;
            ret.put(property.toString(), val);
        }
    } catch (Exception e) {e.printStackTrace();}
    return ret;
}

```

Listato C.1.3

L'inserimento in coda all'indice degli attributi utili per l'ordinamento viene ottenuto mediante il metodo **getIndexingProperties(String)** riportato nel listato C.1.3. Il metodo crea una stringa che contiene un insieme di attributi in base ai quali è possibile ordinare le triple relative ai risultati di una ricerca senza la necessità di dover interrogare la base di conoscenza per ciascuna occorrenza. Ad esempio, il tipo della risorsa (proprietà `rdf:type`), le responsabilità autoriali (`corago:indexingAuthor`) o le date di riferimento (`corago:approxDate`). In questo modo l'applicazione che gestisce la visualizzazione dei risultati può presentare l'elenco secondo l'ordinamento richiesto dall'utente. Con questo approccio è possibile rinviare il caricamento degli ulteriori attributi alla fase di paginazione circoscrivendo l'accesso all'*endpoint* SPARQL ai soli contenuti effettivamente utilizzati dalla vista utente.

Questo 'espediente' implementativo viene riportato a titolo esemplificativo dei diversi che sono stati introdotti per gestire l'accesso ai contenuti. L'ottimizzazione delle strategie di accesso ai dati in formato RDF triple è risultato essere uno dei problemi più significativi nella realizzazione dell'applicazione di consultazione. Questo è un ambito che sicuramente richiede approfondimenti ed evoluzione delle diverse componenti software per poter raggiungere livelli di performance comparabili con le soluzioni basate su basi dati relazionali (vedi paragrafo 5.5).

C.2. Strumenti di analisi NLP

La sperimentazione dell'estrazione di *features* dei personaggi dai testi delle didascalie illustrata nel capitolo 4 si è avvalsa di componenti software di terze parti (vedi paragrafo 4.2). Gli sviluppi più significativi svolti in proprio hanno riguardato gli strumenti per la costruzione dei *training sets*, l'estrazione delle dipendenze sintattiche e la loro rappresentazione in formato RDF. I componenti software sviluppati sono due: il generatore di training set e l'annotatore NLP-RDF.

C.2.1. Generatore training set

La generazione dei *training set* necessari per il riconoscimento di toponimi e dei personaggi come *named entities* (NER) è implementata nel metodo `buildCharacterTS()` della classe `TrainingSetBuidler` (vedi listato C.2.1). Il metodo estrae dalla base di conoscenza sia le etichette (`rdfs:label`) delle risorse da classificare alternativamente come 'PER'=personaggi o 'LOC'=toponimi che i testi delle didascalie ed ambientazioni delle opere attraverso un metodo ausiliario. I contenuti estratti vengono convertiti nel tracciato previsto dallo StanfordCore NLP e serializzati in un file di tipo TSV. Nel nostro caso si è potuto sfruttare una fonte che contiene un elenco controllato di nomi di istanze per entrambe le classi. Questo ha consentito l'annotazione di tutte le occorrenze presenti nei testi indipendentemente dalla loro provenienza (ad esempio un toponimo può essere citato in una didascalia di un personaggio la cui opera non è necessariamente associata ad esso). Il risultato è un *training set* generato automaticamente su un numero significativo di occorrenze¹⁹⁰ ed al tempo stesso poco rumoroso.

¹⁹⁰ Il training set ottenuto contiene oltre 400k annotazioni per una dimensione complessiva di 3.6MB.

```

public static void buildCharacterTSV()
{
    List<String> tsLines = new ArrayList<String>();
    // estrarre le triple da utilizzare come fonte
    List<Triple> charList = SPARQLConnector.getCharacterResourceData();
    // costruire indice dei nomi delle PER
    ArrayList<String> charIndex = new ArrayList<String>();
    Iterator<Triple> ai = charList.iterator();
    while (ai.hasNext()) {
        Triple aT = ai.next();
        if (!charIndex.contains(aT.getSubject().toString()) && aT.getSubject().isLiteral()) //escludo gli URI
            charIndex.add(aT.getSubject().toString().trim());
    }
    // per ogni tripla splittare la qualifica e annotarla con la classe PER (se in indice nomi) oppure O
    // accumulare le annotazioni in formato "word TAB class LINEFEED"
    Iterator<Triple> charIter = charList.iterator();
    int tot = charList.size();
    int cnt = 0;
    while (charIter.hasNext()) {
        cnt++;
        System.out.println("Processo " + cnt + " di " + tot);
        Triple aT = charIter.next();
        String charName = aT.getSubject().toString();
        String charQual = aT.getObject().toString();
        // aggiungo il nome
        charName = charName.replaceAll("\\", "");
        String tsLine = charName.replace(" ", "-") + "\t" + "PER" + "\n";
        tsLines.add(tsLine);
        // splitto e analizzo la qualifica
        charQual = charQual.replaceAll("(", "-");
        charQual = charQual.replaceAll("\\.", "-");
        ArrayList<String> qualChunks = new ArrayList<String>(Arrays.asList(charQual.split(" ")));
        Iterator<String> qualChunksIter = qualChunks.iterator();
        while (qualChunksIter.hasNext())
        {
            String charQualChunk = qualChunksIter.next();
            charQualChunk = charQualChunk.replaceAll("[\n\r]", "");
            charQualChunk = charQualChunk.trim().toString();
            Boolean isPers=false;
            for(String charIndexEntry: charIndex)
            {if (StringUtils.getLevenshteinDistance(charIndexEntry.toString(),charQualChunk.toString())<3) {
                isPers=true;
                break;}}
            if (isPers) { // aggiungo il nome
                tsLine = charQualChunk + "\t" + "PER" + "\n";
                tsLines.add(tsLine);
            } else {
                if(SPARQLConnector.isDBPediaToponym(charQualChunk))
                { // aggiungo il toponimo
                    tsLine = charQualChunk + "\t" + "LOC" + "\n";
                    tsLines.add(tsLine);}
                else if(charQualChunk.trim().length()>0){ // aggiungo il background
                    tsLine = charQualChunk + "\t" + "O" + "\n";
                    tsLines.add(tsLine);}}}}
    }
    // serializzare nel file TSV di output
    File outFile = new File(".\\trainingset.tsv");
    try {
        FileUtils.writeStringToFile(outFile,tsLines.toString().replaceAll(" ", "").replaceAll("[", "").replaceAll("]", ""));
    } catch (FileNotFoundException e) {e.printStackTrace();} catch (IOException e) {e.printStackTrace();}
}

```

Listato C.2.1

C.2.2. Annotatore NLP-RDF

L'obiettivo di questo componente è produrre l'annotazione delle dipendenze morfosintattiche delle didascalie dei personaggi in formato RDF (vedi paragrafo 4.2). Per ottenere questo risultato il metodo `extractDep(Map<String,Triple>)` della classe `DependenciesExtractor` accetta in firma le coppie: URI della risorsa personaggio e la tripla relativa alla sua didascalia. A partire da queste ritorna un elenco di triple che rappresentano il grafo delle dipendenze estratte dal testo della descrizione oltre alle annotazioni PoS e NER. Il metodo è quindi strutturato in tre sezioni: l'annotazione dei *token*, la conversione in RDF dei livelli relativi al PoS e la creazione del grafo relativo alle dipendenze UD.

Il listato C.2.2a riporta la porzione di codice relativo alla fase iniziale di annotazione del testo che impiega la *pipeline* TINT. Va segnalata la necessità di limitare il numero di iterazioni a causa di problemi nella gestione della memoria utilizzata dal processo quando si supera una certa soglia di iterazioni che, nel nostro caso, è stata empiricamente individuata in circa 1.5K (porzioni di codice evidenziate in grassetto).¹⁹¹

¹⁹¹ Non è stato possibile approfondire le cause del comportamento osservato che, a partire dalla soglia individuata, produce un aumento del consumo dello spazio di *heap* pari a 2GB assegnato al processo tale da generare una eccezione di tipo `java.lang.OutOfMemoryError`. Si è quindi adottato un approccio preventivo con la inizializzazione di una nuova istanza della classe `TintPipeline`.

```

public static List<Triple> extractDep(Map<String,Triple> uriTripleMap)
{
    List<Triple> ret = new ArrayList<Triple>();
    //contatore limite memoria: migliorare la gestione dell'eccezione TINT
    Integer cnt =0;
    //definizione namespaces utilizzati
    String udNs = "http://universaldependencies.org/u/dep#";
    String coragoNs = "http://corago.unibo.it/smf/";
    //inizializzo la pipeline
    TintPipeline pipeline = new TintPipeline();
    try {
        pipeline.loadDefaultProperties();
    } catch (IOException e1) {e1.printStackTrace();}
    pipeline.load();
    // itero sulla mappa in ingresso
    for (String aCharURI: uriTripleMap.keySet()) {
        cnt++;
        Triple aTr = uriTripleMap.get(aCharURI);
        //creo la tripla radice: URI personaggio -> hasSentence -> Qualifica URI
        org.apache.jena.graph.Node perUri = NodeFactory.createURI(aCharURI);
        org.apache.jena.graph.Node prop = NodeFactory.createURI(coragoNs + "hasSentence");
        org.apache.jena.graph.Node sntUri = NodeFactory.createURI(aCharURI+"/1");
        Triple baseTr = new Triple(perUri, prop, sntUri);
        ret.add(baseTr);
        //aggiungo la label relativa alla qualifica
        prop = NodeFactory.createURI("http://www.w3.org/2000/01/rdf-schema#label");
        org.apache.jena.graph.Node sntLab = NodeFactory.createLiteral(aTr.getObject().toString().replace("\\", ""));
        Triple qualTr = new Triple(sntUri, prop, sntLab);
        ret.add(qualTr);
        // trasformazione in stringa S+O (concatenazione del formato per Personaggi/Qualifiche)
        String textToParse = (aTr.getSubject().toString() + " " + aTr.getObject().toString()).replaceAll("\\",
        "").replace("[!]", "");
        // annotazione
        InputStream stream = new ByteArrayInputStream(textToParse.getBytes(StandardCharsets.UTF_8));
        ByteArrayOutputStream outXML = new ByteArrayOutputStream();
        Annotation annotation = null;
        try {
            System.out.println(cnt.toString() + " - " + textToParse.toString());
            annotation = pipeline.run(stream, outXML, TintRunner.OutputFormat.XML);
            if(cnt>1500)
            {
                System.out.println("WARN: reinizializzazione della pipeline!");
                pipeline = new TintPipeline();
                try {
                    pipeline.loadDefaultProperties();
                } catch (IOException e1) {e1.printStackTrace();}
                pipeline.load();
                cnt=0;
            }
        } catch (Exception e) {System.out.println(e.getMessage());
            continue;}
    }
}

```

[...]

Listato C.2.2a

[...]

```
InputStream xmlIS = new ByteArrayInputStream(outXML.toByteArray());
DocumentBuilderFactory factory = DocumentBuilderFactory.newInstance();
DocumentBuilder xmlBuild;
XPathFactory xPathfactory = XPathFactory.newInstance();
XPath xpath = xPathfactory.newXPath()
try {
    xmlBuild = factory.newDocumentBuilder();
    InputSource iS = new InputSource(new StringReader(outXML.toString()));
    Document xmlParsed = xmlBuild.parse(iS);
    //estrazione dei TOKENS
    XPathExpression tokensExpr = xpath.compile("/root/document/sentences/sentence[@id='1']/tokens/token");
    String sntId = "1"; // analisi limitata alla sola prima frase
    XPathExpression tkid = xpath.compile("./@id");
    XPathExpression word = xpath.compile("./word");
    XPathExpression lemma = xpath.compile("./lemma");
    XPathExpression POS = xpath.compile("./POS");
    XPathExpression NER = xpath.compile("./NER");
    // preparazione delle Triple RDF dei token
    if(xmlParsed!=null){
        NodeList tokens = (NodeList) tokensExpr.evaluate(xmlParsed, XPathConstants.NODESET);
        if(tokens.getLength()>0){
            for(int i=0;i<tokens.getLength();i++){
                String tknId = (String) tkid.evaluate(tokens.item(i), XPathConstants.STRING);
                //creo la relazione tra sentence e token (perUri+sentenceld+tokenld)
                org.apache.jena.graph.Node tokenUri = NodeFactory.createURI(perUri.toString()+"/"+sntId+"/"+tknId);
                org.apache.jena.graph.Node pred = NodeFactory.createURI(udNs+"hasToken");
                Triple depRelRDF = new Triple(sntUri, pred, tokenUri);
                ret.add(depRelRDF);
                //creo il tipo Token
                pred = NodeFactory.createURI("http://www.w3.org/1999/02/22-rdf-syntax-ns#type");
                org.apache.jena.graph.Node type = NodeFactory.createURI(udNs+"TOKEN");
                depRelRDF = new Triple(tokenUri, pred, type);
                ret.add(depRelRDF);
                //creo tripla per la word
                pred = NodeFactory.createURI(udNs+"word");
                String tknWord = ((org.w3c.dom.Node)word.evaluate(tokens.item(i),XPathConstants.NODE)).getTextContent();
                org.apache.jena.graph.Node obj = NodeFactory.createLiteral(tknWord);
                depRelRDF = new Triple(tokenUri, pred, obj);
                ret.add(depRelRDF);
                //creo tripla per il lemma
                pred = NodeFactory.createURI(udNs+"lemma");
                obj =
                NodeFactory.createLiteral(((org.w3c.dom.Node)lemma.evaluate(tokens.item(i),XPathConstants.NODE)).getTextContent());
                depRelRDF = new Triple(tokenUri, pred, obj);
                ret.add(depRelRDF);
                //creo tripla per il PoS
                pred = NodeFactory.createURI(udNs+"POS");
                obj =
                NodeFactory.createLiteral(((org.w3c.dom.Node)POS.evaluate(tokens.item(i),XPathConstants.NODE)).getTextContent());
                depRelRDF = new Triple(tokenUri, pred, obj);
                ret.add(depRelRDF);
                //creo tripla per la NER
                pred = NodeFactory.createURI(udNs+"NER");
                String tknNER = ((org.w3c.dom.Node)NER.evaluate(tokens.item(i),XPathConstants.NODE)).getTextContent();
                obj = NodeFactory.createLiteral(tknNER);
                depRelRDF = new Triple(tokenUri, pred, obj);
                ret.add(depRelRDF);
                //aggancio il token all'uriPer se coincidono etichetta e NER
                if(tknNER.equals("PER") && aTr.getSubject().toString().contains(tknWord)) {
```

Il listato C.2.2b riporta la sezione del metodo che produce le triple relative alle annotazioni dei *tokens* estratti dalla sezione precedente a livello di ‘parola’ (word), lemma, PoS e NER. Vengono create le triple relative e la relazione diretta con il personaggio attraverso la specifica proprietà corago:CNLP1_refers_to_character.

```
[...]
//analisi della struttura delle dipendenze
XPathExpression dependencies = xpath
.compile("/root/document/sentences/sentence/dependencies[@type='basic-dependencies']/dep");
XPathExpression depType = xpath.compile("./@type");
XPathExpression governor = xpath.compile("./governor/@idx");
XPathExpression dependent = xpath.compile("./dependent/@idx");
// elenco delle PER individuate
NodeList depPerXML = (NodeList) dependencies.evaluate(xmlParsed, XPathConstants.NODESET);
if (depPerXML.getLength() > 0) {
    for (int i = 0; i < depPerXML.getLength(); i++) {
        org.apache.jena.graph.Node sbj = NodeFactory.createURI(
            (sntUri.toString() + "/" + (String) governor.evaluate(depPerXML.item(i), XPathConstants.STRING)));
        org.apache.jena.graph.Node pred = NodeFactory
            .createURI(udNs + (String) depType.evaluate(depPerXML.item(i), XPathConstants.STRING));
        org.apache.jena.graph.Node obj = NodeFactory.createURI(
            (sntUri.toString() + "/" + (String) dependent.evaluate(depPerXML.item(i), XPathConstants.STRING)));
        Triple depRelRDF = new Triple(sbj, pred, obj);
        ret.add(depRelRDF);
    }
}
} catch (ParserConfigurationException e) {e.printStackTrace();}
} catch (SAXException e) {e.printStackTrace();}
} catch (IOException e) {e.printStackTrace();}
} catch (XPathExpressionException e) {e.printStackTrace();}
}
}
return ret;
}
```

Listato C.2.2c

Il listato C.2.2c riporta la sezione del metodo dedicata alla estrazione delle UD tra i *tokens* e la relativa rappresentazione come triple RDF. Questo è il passaggio finale che ha prodotto il *dataset* impiegato per estrarre le relazioni interpersonali dei personaggi mediante attraverso la costruzione di regole basate su *pattern* sintattici espressi in linguaggio SPARQL (vedi paragrafo 4.2.4).

Gli esempi di codice illustrano il processo di integrazione dei risultati prodotti da strumenti software non nativamente in grado di interagire con risorse in formato RDF in un contesto interamente basato su linguaggi e tecnologie semantiche.

C.3. Utilizzo delle proprietà di navigazione

Nel paragrafo 2.5 è stato presentato il modello di navigazione dei contenuti basato sulla semantica di dominio. Lo strumento utilizzato per formalizzare la correlazione tra quest'ultima e le triple nel *dataset* in base al modello concettuale derivato dalle due ontologie di riferimento, sono le cosiddette 'proprietà di navigazione' ('navProp'). Queste vengono definite come OWL Object Properties all'interno dell'ontologia di dominio Corago SM. Attraverso la condivisione dell'*ancestor* comune (la proprietà astratta *corago:CPA2_navigable*) queste costituiscono il *layer* che l'applicazione di consultazione impiega per rappresentare le risorse nelle viste utente. La relazione tra proprietà e la corrispondente porzione di grafo è specificata come *property path* della proprietà di navigazione. Per produrre una rappresentazione della risorsa in base alle 'navProps', l'applicazione di consultazione dovrà semplicemente individuare quelle riferite al tipo della risorsa in oggetto e tradurre la *property path* di ciascuna in una interrogazione SPARQL per estrarre i contenuti corrispondenti.

```

import java.util.ArrayList;
import java.util.HashMap;
import java.util.List;
import java.util.Map;

public class NavPropertyBean implements Comparable<NavPropertyBean> {
    private String propIri, propChain;
    private int priority, typingClassPosition;
    private Map<String,String> propLabels, propComments;
    private String typingClass,thematicArea;
    private List<String> domainClasses, rangeClasses;
    private boolean isFeature=false;

    public NavPropertyBean()
    {
        propLabels = new HashMap<String,String>();
        propComments = new HashMap<String,String>();
    }

    public String getPropIri() {
        return propIri;
    }
    public void setPropIri(String propIri) {
        this.propIri = propIri;
    }
    public Map<String, String> getPropLabels() {
        return propLabels;
    }
    public void setPropLabels(Map<String, String> propLabels) {
        this.propLabels = propLabels;
    }
    public Map<String, String> getPropComments() {
        return propComments;
    }
    public void setPropComments(Map<String, String> propComments) {
        this.propComments = propComments;
    }
    @Override
    public String toString() {
        return propIri;
    }
    public String getPropChain() {
        return propChain;
    }
    public void setPropChain(String propChain) {
        this.propChain = propChain;
    }
    public void addLabel(String label, String lang)
    {
        propLabels.put(lang, label);
    }
    public void addComment(String comment, String lang)
    {
        propComments.put(lang, comment);
    }
}

```

[...]

Listato C.3.1a

```

[...]  

    public String getLabel(String lang)  

    {  

        return propLabels.get(lang);  

    }  

    public String getComment(String lang)  

    {  

        return propComments.get(lang);  

    }  

    public int getPriority() {  

        return priority;  

    }  

    public void setPriority(int priority) {  

        this.priority = priority;  

    }  

    @Override  

    public int compareTo(NavPropertyBean o) {  

        if (this.getPriority()<o.getPriority()){  

            return -1;  

        }else{  

            return 1; }  

    }  

    public int getTypingClassPosition() {  

        return typingClassPosition;  

    }  

    public void setTypingClassPosition(int typingClassPosition) {  

        this.typingClassPosition = typingClassPosition;  

    }  

    public String getTypingClass() {  

        return typingClass;  

    }  

    public void setTypingClass(String typingClass) {  

        this.typingClass = typingClass;  

    }  

    public String getThematicArea() {  

        return thematicArea;  

    }  

    public void setThematicArea(String thematicArea) {  

        this.thematicArea = thematicArea;  

    }  

    public List<String> getDomainClasses() {  

        return domainClasses;  

    }  

    public void setDomainClasses(List<String> domainClasses) {  

        this.domainClasses = domainClasses;  

    }  

    public List<String> getRangeClasses() {  

        return rangeClasses;  

    }  

    public void setRangeClasses(List<String> rangeClasses) {  

        this.rangeClasses = rangeClasses;  

    }  

    public boolean isFeature() {  

        return isFeature;  

    }  

    public void setIsFeature(boolean isFeature) {  

        this.isFeature = isFeature;  

    }  

}

```

Listato C.3.1b

Le 'navProp' sono implementate nell'applicazione sperimentale attraverso una classe specifica "NavPropertyBean" di cui viene creata una istanza in fase di inizializzazione per ciascuna proprietà di navigazione prevista dal Corago SM. Il listato C.3.1 (parti *a* e *b*) riporta il codice relativo alla definizione della classe. Si noti la proprietà **PropChain** con i relativi metodi *getter* e *setter* utilizzata per conservare la stringa relativa alla *property path* corrispondente alla specifica istanza di 'navProp'. L'inizializzazione della collezione di proprietà di navigazione in forma di Java *beans* avviene in fase di avvio dell'applicazione ad opera del metodo: `populateNavigablePC(OWLOntology corago)` della classe `ConfigurationBean` adibita alla raccolta delle configurazioni di sistema.

Il listato C.3.2 (parti *a* e *b*) illustra invece la funzione che, in fase di inizializzazione, provvede a leggere l'ontologia Corago SM e ad istanziare le proprietà di navigazione previste per ciascuna tipologia di risorsa.

```

public void populateNavigablePC(OWLOntology corago) {
    // gestisce la valorizzazione della configurazione delle proprietà di navigazione
    OWLObjectProperty objPropAncestor = corago.getOWLOntologyManager().getOWLDataFactory()
        .getOWLObjectProperty(IRI.create(this.getNavigableRelationAncestor()));
    OWLDataProperty dataPropAncestor = corago.getOWLOntologyManager().getOWLDataFactory()
        .getOWLDataProperty(IRI.create("http://corago.unibo.it/sm/CDA2_navigable"));
    Set<OWLSubObjectPropertyOfAxiom> objProps = corago.getObjectSubPropertyAxiomsForSuperProperty(objPropAncestor);
    Set<OWLSubDataPropertyOfAxiom> dataProps = corago.getDataSubPropertyAxiomsForSuperProperty(dataPropAncestor);
    Set<OWLSubPropertyChainOfAxiom> propChains = corago.getAxioms(AxiomType.SUB_PROPERTY_CHAIN_OF);

    // itero sulle navigation object properties dell'ontologia
    for (OWLSubObjectPropertyOfAxiom aSub : objProps) {
        String sprqlClause = "";
        String pcLabel = "";
        int priority = 0, typingClassPosition = 0;
        ArrayList<String> pcForProperty = new ArrayList<String>();
        Iterator<OWLSubPropertyChainOfAxiom> ai = propChains.iterator();
        // itero sulle proprietà navigabili
        while (ai.hasNext()) {
            OWLSubPropertyChainOfAxiom aPC = ai.next();
            // se trovata come super proprietà allora estraggo la property chain e costruisco l'equivalente SPARQL
            if (aSub.getSubProperty().toString().equals(aPC.getSuperProperty().asOWLObjectProperty().toString())) {
                String pcClause = "";
                Iterator<OWLObjectPropertyExpression> pcl = aPC.getPropertyChain().iterator();
                while (pcl.hasNext()) {
                    OWLObjectPropertyExpression pcE = pcl.next();
                    if (pcClause.length() > 0)
                        pcClause += "/";
                    if (!pcE.isAnonymous())
                        pcClause += Misc.removePrefix(pcE.asOWLObjectProperty().getIRI().toString(), this);
                    else
                        pcClause += Misc.removePrefix(
                            "^" + pcE.getSimplified().getNamedProperty().getIRI().toString(), this);
                    pcForProperty.add(pcClause);
                }
                // concateno la stringa SPARQL sommando in OR ciascuna property chain
                Iterator<String> pclt = pcForProperty.iterator();
                while (pclt.hasNext()) {
                    String clToAdd = pclt.next();
                    sprqlClause += clToAdd;
                    if (pclt.hasNext())
                        sprqlClause += "|";
                }
                // valorizzo le proprietà in base alla definizione della property
                pcLabel = Misc.getAnnotationValue(aSub.getSubProperty().asOWLObjectProperty(),
                    "http://www.w3.org/2000/01/rdf-schema#label", this.getPreferredLanguage(), corago);
                String pri = Misc.getAnnotationValue(aSub.getSubProperty().asOWLObjectProperty().getIRI(),
                    "http://corago.unibo.it/sm/CPAN1_priority", this.getPreferredLanguage(), corago);
                String isFeat = Misc.getAnnotationValue(aSub.getSubProperty().asOWLObjectProperty().getIRI(),
                    "http://corago.unibo.it/sm/CPAN5_is_feature", this.getPreferredLanguage(), corago);
                String typingClass = Misc.getAnnotationValue(aSub.getSubProperty().asOWLObjectProperty().getIRI(),
                    "http://corago.unibo.it/sm/CPAN2_typing_class", this.getPreferredLanguage(), corago);
                String typingClassPos = Misc.getAnnotationValue(aSub.getSubProperty().asOWLObjectProperty().getIRI(),
                    "http://corago.unibo.it/sm/CPAN3_typing_class_position", this.getPreferredLanguage(), corago);
                String thematicArea = Misc.getAnnotationValue(aSub.getSubProperty().asOWLObjectProperty().getIRI(),
                    "http://corago.unibo.it/sm/CPAN4_thematic_area", this.getPreferredLanguage(), corago);
            }
        }
    }
}

```

[...]

Listato C.3.2a

[...]

```
// valorizzo domain e range della property
Set<OWLObjectPropertyDomainAxiom> domains = corago.getObjectPropertyDomainAxioms(aSub.getSubProperty());
Set<OWLObjectPropertyRangeAxiom> ranges = corago.getObjectPropertyRangeAxioms(aSub.getSubProperty());
Set<OWLSubObjectPropertyOfAxiom> subs = corago
    .getObjectSubPropertyAxiomsForSubProperty(aSub.getSubProperty());

if (pri != "n.d.")
    priority = Integer.parseInt(Misc.stripValue(pri));
if (typingClassPos != "n.d.")
    typingClassPosition = Integer.parseInt(typingClassPos);

NavPropertyBean aNb = new NavPropertyBean();
if (domains != null && !domains.isEmpty()) {
    Iterator<OWLObjectPropertyDomainAxiom> aid = domains.iterator();
    ArrayList<String> domainClasses = new ArrayList<>();
    while (aid.hasNext()) {
        domainClasses.add(aid.next().getDomain().toString());
    }
    aNb.setDomainClasses(domainClasses);
}
if (ranges != null && !ranges.isEmpty()) {
    Iterator<OWLObjectPropertyRangeAxiom> air = ranges.iterator();
    ArrayList<String> rangeClasses = new ArrayList<>();
    while (air.hasNext()) {
        rangeClasses.add(air.next().getRange().toString());
    }
    aNb.setRangeClasses(rangeClasses);
}
aNb.setPropChain(sprqlClause);
aNb.setPropIri(aSub.getSubProperty().asOWLObjectProperty().getIRI().toString());
aNb.addLabel(pcLabel, "it");
aNb.setPriority(priority);
aNb.setTypingClass(typingClass);
aNb.setTypingClassPosition(typingClassPosition);
aNb.setThematicArea(thematicArea);
if (isFeat != null && isFeat.equals("true"))
    aNb.setIsFeature(true);

// estraggo il dominio
OWLObjectProperty oP = aSub.getSubProperty().asOWLObjectProperty();
Set<OWLObjectPropertyDomainAxiom> propDom = corago.getAxioms(AxiomType.OBJECT_PROPERTY_DOMAIN);
Iterator<OWLObjectPropertyDomainAxiom> apDI = propDom.iterator();
while (apDI.hasNext()) {
    OWLObjectPropertyDomainAxiom apD = apDI.next();
    // aggancio solo se il dominio è ammesso e per tutte le eventuali sottoclassi
    if (apD.getProperty().toString().equals(oP.toString())) {
        this.addCoragoNavigationCRProp(apD.getDomain().toString(), aNb);
        //valuto le sottoclassi del dominio
        List<OWLSubClassOfAxiom> subPropDomains = new ArrayList<OWLSubClassOfAxiom>();
        for (final OWLSubClassOfAxiom subClasse : corago.getAxioms(AxiomType.SUBCLASS_OF))
        {
            if (subClasse.getSuperClass().getIRI().equals(apD.getDomain().getIRI())) {
                subPropDomains.add(subClasse);
                System.out.println(subClasse.getSubClass().getIRI() + " extends " + subClasse.getSuperClass().getIRI());
            }
        }
        for(OWLSubClassOfAxiom aSubDom:subPropDomains)
        {
            this.addCoragoNavigationCRProp(aSubDom.getSubClass().getIRI().toString(), aNb);
        }
    }
}
}
```

Listato C.3.2b

Il listato evidenzia come l'unico parametro utilizzato dal metodo di inizializzazione della configurazione che guida tutta la logica di costruzione delle viste utente sia l'ontologia di dominio (rappresentata all'interno della soluzione come oggetto di tipo `org.semanticweb.owlapi.model.OWLOntology`¹⁹²). L'implementazione illustra come realizzare delle rappresentazioni dei contenuti efficaci dal punto di vista funzionale (vedi paragrafo 5.5) a partire dal solo modello concettuale del dominio espresso attraverso un'ontologia in formato OWL.

Il recupero dei contenuti delle proprietà avviene attraverso la costruzione di una query SPARQL basata sulla trasformazione della *property chain* dalla sintassi OWL alla relativa sintassi SPARQL. La composizione dell'interrogazione in fase di estrazione dei dati è delegata nell'applicazione sperimentale ad un '*query builder*'. Il listato C.3.3 riporta la versione più generica tra i diversi metodi implementati a questo scopo.

```
public String buildNavPropQuery(NavPropertyBean property, TripleBean triple, ConfigurationBean conf)
{
    String ret = "";
    String classIri = Misc.classToIri(triple.getType(), conf).toString();
    String clauseQuery = "{";
    currentQuery = prefixQuery + "SELECT distinct (<" + triple.getIri() + "> as ?s) (<" + property.getPropIri() + "> as ?p) ?o "
    clauseQuery = clauseQuery + "<" + triple.getIri() + "> " + property.getPropChain() + " ?o";
    if(clauseQuery!="{")
        currentQuery = currentQuery + clauseQuery + "}";
    //ritorno la query relativa alla proprietà navigabile
    ret=currentQuery;
    return ret;
}
```

Listato C.3.3

Si può osservare come il metodo consumi direttamente la proprietà contenente la serializzazione in formato SPARQL della *property chain* OWL relativa all'istanza di `NavPropertyBean` passata in firma attraverso il metodo `getPropChain()`. Nell'architettura di sistema adottata per l'applicazione sperimentale si è quindi scelto di isolare le componenti di modello utilizzando lo strato di oggetti (*beans*) di configurazione come *adapter* tra i due mondi. Questa resta una scelta di natura puramente implementativa. Infatti, una applicazione terza che volesse utilizzare i contenuti della base di conoscenza Corago seguendo la semantica di dominio definita nel Corago SM potrà adottare soluzioni architetturali differenti.

¹⁹² La libreria utilizzata per la gestione delle ontologie in formato OWL è OWL API la cui documentazione è reperibile all'indirizzo: <http://owlcs.github.io/owlapi/> (verificato il 23/08/2019).

Bibliografia

- Accorsi, M.G., 1989. Problemi testuali dei libretti d'opera fra Sei e Settecento. *Giornale storico della letteratura italiana* CLXVI.
- Alexiev, V., 2012. Types and Annotations for CIDOC CRM Properties. Presented at the Digital Presentation and Preservation of Cultural and Scientific Heritage (DiPP2012), Springer Verlag, Berlin-Heidelberg.
- Allenmang, D., Hendler, J., 2011. Semantic web for the working ontologist. Morgan Kaufman.
- Aproso, A.P., Moretti, G., 2016. Italy goes to Stanford: a collection of CoreNLP modules for Italian. *arXiv:1609.06204 [cs]*.
- Bartolini, I., Luzzi, C., Patella, M., Pompilio, A., 2012. MIMESIS: A Semantic-based Approach for a Digital Library of, in: Poetry in Music. in *Proceedings ECLAP 2012: Conference on Information Technologies for Performing Arts. Media Access and Entertainment: Florence 7-9*. Firenze University Press.
- Bassett, C., Berry, D.M., Fazi, M.B., Pay, J., Roberts, B., 2017. Critical Digital Humanities and Machine Learning. DH.
- Bast, H., Bäurle, F., Buchhold, B., Haussmann, E., 2012. A case for semantic full-text search, in: *Proceedings of the 1st Joint International Workshop on Entity-Oriented and Semantic Search - JIWES '12*. Presented at the the 1st Joint International Workshop, ACM Press, Portland, Oregon, pp. 1–3. <https://doi.org/10.1145/2379307.2379311>
- Bast, H., Buchhold, B., 2013. An index for efficient semantic full-text search, in: *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management - CIKM '13*. Presented at the the 22nd ACM international conference, ACM Press, San Francisco, California, USA, pp. 369–378. <https://doi.org/10.1145/2505515.2505689>
- Bekiari, Chrysoula, Doerr, M., Le Boef, P., Riva, P., 2015. FRBR object-oriented definition and mapping from FRBRer, FRAD and FRSAD. *ICOM-CIDOC*.
- Bellini, P., Nesi, P., 2014. Modeling Performing Arts Metadata and Relationships in Content Service for Institutions. *Multimedia Systems*. Springer, Berlin-Heidelberg. <https://doi.org/10.1007/s00530-014-0366-0>
- Berardi, G., Esuli, A., Marcheggiani, D., 2015. Word Embeddings Go to Italy: a Comparison of Models and Training Datasets. *IIR*.
- Bianconi, L., 2017. Il libretto d'opera, in: *Musica*. Istituto della Enciclopedia Italiana.
- Bianconi, L., Pompilio, A., Pagannone, G., 2004. RADAMES: prototipo d'un repertorio e archivio digitale per il melodramma. *Il Saggiatore Musicale*.
- Bignami, P., Azzaroni, G., 1997. Gli oggetti nello spazio del teatro: atti dei Convegni Spazi del teatro, idee e luoghi di spettacolo e Il teatro degli oggetti, gli oggetti del teatro. Bulzoni.
- Bizer, C., Cyganiak, R., Heath, T., 2007. How to publish linked data on the web.
- Bizer, C., Heath, T., Berners-Lee, T., 2011. Linked Data: The story so far, in: *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. IGI Global, pp. 205–227.
- Bizer, C., Seaborne, A., 2004. D2RQ-treating non-RDF databases as virtual RDF graphs, in: *Proceedings of the 3rd International Semantic Web Conference (ISWC2004)*. *Proceedings of ISWC2004*.
- Bonlini, C., 1730. *Le glorie della poesia e della musica*. Venezia.
- Bonora, P., 2016. Trasposizione di un indice repertoriale del melodramma nel paradigma Linked Open Data (Metodologie informatiche per le Discipline Umanistiche). Università di Firenze, Firenze.
- Bonora, P., Ossicini, C., Raffa, G., 2006. From Relational Metadata Standards to CRM Ontology: a Case Study in Performing Arts, in: *CIDOC2006 Proceedings*. Presented at the CIDOC2006, Gothenburg.
- Bonora, P., Pompilio, A., 2019. "Osservate, leggete con me". Risorse LOD per la storia del melodramma: una prospettiva funzionale di rappresentazione. Presented at the 8th Annual Conference AIUCD 2019, Udine.
- Börner, K., Conlon, M., Corson-Rikert, J., Ding, Y. (Eds.), 2012. VIVO: a semantic approach to scholarly networking and discovery, *Synthesis lectures on the semantic web: theory and technology*. Morgan & Claypool, S.I.

- Calhoun, K., 2011. The Changing Nature of the Catalog and its Integration with other Discovery Tools, in: McIntosh, J. (Ed.), *Cataloging and Indexing*. Apple Academic Press, pp. 123–173. <https://doi.org/10.1201/b13123-9>
- Calvanese, D., Mosca, A., Remesal, J., Rezk, M., Rull, G., 2015. A ‘historical case’ of ontology-based data access, in: *2015 Digital Heritage*. IEEE, pp. 291–298.
- Carriero, V.A., Daquino, M., Tomasi, F., 2019. Semantic alignment in museums, archives and libraries. The ontologies for describing relationships. *JLIS. it* 10, 72–91.
- Chen, Y., Zhao, X., Zhang, S., 2013. Publishing rdf from relational database based on d2R improvement. *WSEAS Transactions on Information Science and Applications* 10, 241–248.
- Ciotti, F., Tomasi, F., 2016. Formal ontologies, linked data, and TEI semantics. *Journal of the Text Encoding Initiative*.
- Cipollone, B., 2009. La catalogazione dei libretti nel sistema di archiviazione digitale Il Corago. *Quaderni Estensi Rivista* 1.
- Colazzo, D., Goasdoué, F., Manolescu, I., Roatiş, A., 2014. RDF analytics: lenses over semantic graphs, in: *Proceedings of the 23rd International Conference on World Wide Web - WWW '14*. Presented at the the 23rd international conference, ACM Press, Seoul, Korea, pp. 467–478. <https://doi.org/10.1145/2566486.2567982>
- Console, M., Lenzerini, M., 2014. Data Quality in Ontology-Based Data Access: the Case of Consistency, in: *Twenty-Eighth AAAI Conference on Artificial Intelligence*.
- Corbetta, P., 2014. *Metodologia e tecniche della ricerca sociale*. Il Mulino, Bologna.
- Cotticelli, F., Maione, P., 2016. *Opera buffa Napoli 1707-1750*.
- Crawford, T., Fields, B., Lewis, D., Page, K., 2014. Explorations in Linked Data practice for early music corpora, in: *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*. IEEE Press.
- Cruciani, F., 1993. *Lo spazio del teatro*. Laterza, Bari.
- Cruciani, F., 1987. Il teatro e la festa, in: *Il Teatro Italiano Nel Rinascimento*. Il Mulino, Bologna, pp. 31–52.
- Daquino, M., Tomasi, F., 2014. Ontological approaches to information description and extraction in the cultural heritage domain, in: *Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem*. ACM, p. 8.
- De Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., Manning, C., 2014. Universal Stanford Dependencies: A cross-linguistic typology. *LREC* 14.
- De Marneffe, M.-C., Manning, C.D., 2008. *Stanford typed dependencies manual*. Stanford University.
- Ding, Y., Sun, Y., Chen, B., Borner, K., Ding, L., Wild, D., Wu, M., DiFranzo, D., Fuenzalida, A.G., Li, D., Milojevic, S., Chen, S., Sankaranarayanan, M., Toma, I., 2010. Semantic Web Portal: A Platform for Better Browsing and Visualizing Semantic Data, in: An, A., Lingras, P., Petty, S., Huang, R. (Eds.), *Active Media Technology*. Springer, Berlin-Heidelberg, pp. 448–460. https://doi.org/10.1007/978-3-642-15470-6_46
- Doerr, M., 2009. Ontologies for cultural heritage, in: *Ontologies for Cultural Heritage*. Springer Verlag, Berlin-Heidelberg, pp. 463–486.
- Doerr, M., 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AAAI. AI Magazine* 24. <https://doi.org/DOL:http://dx.doi.org/10.1609/aimag.v24i3.1720>
- Doerr, M., Bekiari, C., LeBoeuf, P., 2008. FRBROO a Conceptual Model for Performing Arts. Presented at the 2008 Annual Conference of CIDOC, Athens, pp. 15–18.
- Doerr, M., Gradmann, S., LeBoeuf, P., Aalberg, T., Bailly, R., Olensky, M., 2013. Final Report on EDM – FRBRoo Application Profile Task Force. Europeana. Report, Europeana Project.
- Doerr, M., Le Boef, P., 2007. Modelling Intellectual Processes: The FRBR-CRM Harmonization, in: *Digital Libraries: Research & Development*. First International DELOS Conference. Springer Verlag, Berlin-Heidelberg, pp. 114–123.
- Doty, C., 2013. The Difficulty of An Ontology of Live Performance. *InterActions. UCLA Journal of Education and Information Studies* 9.
- Elam, K., 1988. *Semiotica del teatro*. Il Mulino, Bologna.

- Elbassuoni, S., Ramanath, M., Schenkel, R., Weikum, G., 2010. Searching RDF Graphs with SPARQL and Keywords. *IEEE Data Eng. Bull.* 33, 16–24.
- Euzenat, J., Shvaiko, P., 2007. *Ontology matching*. Springer, Berlin-Heidelberg.
- Eva Mayr, Florian Windhager, 2018. Once upon a Spacetime: Visual Storytelling in Cognitive and Geotemporal Information Spaces. *ISPRS International Journal of Geo-Information* 7, 96. <https://doi.org/10.3390/ijgi7030096>
- Frixione, M., Lieto, A., 2011. Representing and reasoning on typicality in formal ontologies. *ACM Press*, pp. 119–125. <https://doi.org/10.1145/2063518.2063534>
- Gagliardi, F., 2008. A prototype-exemplars hybrid cognitive model of “phenomenon of typicality” in categorization: A case study in biological classification. *Proc. 30th Annual Conf. of the Cognitive Science Society*, Austin, TX 1176–1181.
- Gamallo, P., Garcia, M., Fernandez-Lanza, S., 2012. Dependency-based open information extraction. *Proceedings of the joint workshop on unsupervised and semi-supervised learning in NLP* 10–18.
- Gigliozzi, G., 2003. *Introduzione all’uso del computer negli studi letterari*. Bruno Mondadori.
- Groppa, A., 1745. *Catalogo di tutti i drammi per musica recitati ne’ teatri di Venezia dall’anno 1637 ... fin all’anno presente 1745*. Venezia.
- Guerrini, M., Possemato, T., 2015. *Linked data per biblioteche. archive e musei*. Editrice Bibliografica.
- Hai Dong, Hussain, F.K., Chang, E., 2008. A survey in semantic search technologies, in: 2008 2nd IEEE International Conference on Digital Ecosystems and Technologies. Presented at the 2008 2nd IEEE International Conference on Digital Ecosystems and Technologies (DEST), IEEE, Phitsanuloke, Thailand, pp. 403–408. <https://doi.org/10.1109/DEST.2008.4635202>
- Haklay, M., Weber, P., 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7, 12–18.
- Halpin, H., Hayes, P.J., McCusker, J.P., McGuinness, D.L., Thompson, H.S., 2010. , in: *When Owl:SameAs Isn’t the Same: An Analysis of Identity in Linked Data*. Presented at the In International semantic web conference, Springer, Berlin, Heidelberg, pp. 305–320.
- Harbelot, B., Arenas, H., Cruz, C., 2015. LC3: A spatio-temporal and semantic model for knowledge discovery from geospatial datasets. *Web Semantics: Science, Services and Agents on the World Wide Web* 35, 3–24.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *Prototype Methods and Nearest-Neighbors*, in: *The Elements of Statistical Learning*. Springer New York, New York, NY, pp. 1–25. https://doi.org/10.1007/b94608_13
- He, Z., Deng, S., Xu, X., 2006. Approximation algorithms for k-modes clustering, in: *International Conference on Intelligent Computing*. Springer, Berlin-Heidelberg, pp. 296–302.
- Hevner, A.R., March, S.T., Park, J., Ram, S., 2004. *Design Science in Information Systems Research*. *MIS Quarterly* 28, 75–105.
- Hill, J.W., 2001. *Pescetti, Giovanni Battista*. Grove Music Online.
- Hofmann, T., Scholkopf, B., Smola, A.J., 2006. *A Review of Kernel Methods in Machine Learning*. Mac-Planck-Institute Technical Report 156.
- IFLA Study Group, 1998. *Functional Requirements for Bibliographic Records*, IFLA Section on Cataloguing. K. G. Saur, München.
- Jandelli, C., 2002. *I ruoli nel teatro italiano tra Otto e Novecento. Le lettere*, Firenze.
- Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M., Bizer, C., Lee, R., 2009. Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections, in: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (Eds.), *The Semantic Web: Research and Applications*. Springer, Berlin-Heidelberg, pp. 723–737.
- Lausen, H., Ding, Y., Stollberg, M., Fensel, D., Lara Hernández, R., Han, S., 2005. Semantic web portals: state-of-the-art survey. *Journal of Knowledge Management* 9, 40–49. <https://doi.org/10.1108/13673270510622447>

- Le Boef, P., 2012. Towards Performing Arts Information As Linked Data? Presented at the SIBMAS 2012 Conference: Best Practice Innovative Techniques for Performing Arts Collections. Libraries and Museums, France.
- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C., 2012. DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia 29.
- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., others, 2015. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 167–195.
- Lenci, A., Montemagni, S., Pirrelli, V., 2005. *Testo e computer*. Carocci, Roma.
- Libby, D., McClymonds, M.P., 2001. Marinelli, Gaetano. *Grove Music Online*.
- Lieto, A., 2013. Un approccio cognitivo alla rappresentazione dei concetti nelle ontologie formali. *Sistemi intelligenti* 25, 211–228.
- Luo, Y., Picalausa, F., Fletcher, G.H.L., Hidders, J., Vansummeren, S., 2012. Storing and Indexing Massive RDF Datasets, in: De Virgilio, R., Guerra, F., Velegakis, Y. (Eds.), *Semantic Search over the Web*. Springer, Berlin-Heidelberg, pp. 31–60. https://doi.org/10.1007/978-3-642-25008-8_2
- Lutz, M., Kolas, D., 2007. Rule-based Discovery in Spatial Data Infrastructures. *Transactions in GIS* 11, 317–336.
- Macchia, G., 1966. *Vita avventure e morte di Don Giovanni*. Editori Laterza, Bari.
- Marie, N., Gandon, F., 2014. Survey of linked data based exploration systems, in: *IESD 2014 - Intelligent Exploitation of Semantic Data*. Riva Del Garda, Italy.
- Martin, J.H., Jurafsky, D., 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*, Prentice Hall series in artificial intelligence. Prentice Hall, Pearson Education International, Upper Saddle River, NJ.
- McBride, B., 2004. The Resource Description Framework (RDF) and its Vocabulary Description Language RDFS, in: Staab, S., Studer, R. (Eds.), *Handbook on Ontologies*. Springer, Berlin-Heidelberg, pp. 51–65. https://doi.org/10.1007/978-3-540-24750-0_3
- Meroño-Peñuela, A., Ashkpour, A., Van Erp, M., Mandemakers, K., Breure, L., Scharnhorst, A., Schlobach, S., Van Harmelen, F., 2015. Semantic technologies for historical research: A survey. *Semantic Web* 6, 539–564.
- Metitieri, F., Ridi, R., 2003. *Biblioteche in rete: istruzioni per l'uso*. Editori Laterza, Roma.
- Michel, F., Montagnat, J., Zucker, C.F., 2014. A survey of RDB to RDF translation approaches and tools. *Semantic Web journal* 3.
- Mintz, M., Bills, S., Snow, R., Jurafsky, D., 2009. Distant supervision for relation extraction without labeled data, in: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - ACL-IJCNLP '09*. Presented at the the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference, Association for Computational Linguistics, Suntec, Singapore, p. 1003. <https://doi.org/10.3115/1690219.1690287>
- Moro, A., Li, H., Krause, S., Xu, F., Navigli, R., Uszkoreit, H., 2013. Semantic Rule Filtering for Web-Scale Relation Extraction, in: Salinesi, C., Norrie, M.C., Pastor, Ó. (Eds.), *Advanced Information Systems Engineering*. Springer, Berlin-Heidelberg, pp. 347–362. https://doi.org/10.1007/978-3-642-41335-3_22
- Navigli, R., Ponzetto, S.P., 2012. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence* 193, 217–250.
- Pattueli, M.C., 2018. From uniform identifiers to graphs, from individuals to communities: what we talk about when we talk about linked person data, in: *Challenges and Opportunities for Knowledge Organization in the Digital Age*. Ergon-Verlag, pp. 571–580.
- Pattueli, M.C., 2011. Modeling a domain ontology for cultural heritage resources: A user-centered approach. *Journal of the American Society for Information Science and Technology* 62.

- Peroni, S., Shotton, D., 2012. FaBiO and CiTO: ontologies for describing bibliographic resources and citations. In *Journal of Web Semantics: Science. Services and Agents on the World Wide Web*. Web Semantics: Science, Services and Agents on the World Wide Web 33–43.
- Peroni, S., Tomasi, F., Vitali, F., Zingoni, J., 2014. Semantic Lenses as Exploration Method for Scholarly Articles, in: Catarci, T., Ferro, N., Poggi, A. (Eds.), *Bridging Between Cultural Heritage Institutions*. Springer, Berlin-Heidelberg, pp. 118–129. https://doi.org/10.1007/978-3-642-54347-0_13
- Peuquet, D.J., 1999. Making space for time: issues in space-time data representation, in: *Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99*. Presented at the Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99, IEEE, Florence, Italy, pp. 404–408. <https://doi.org/10.1109/DEXA.1999.795200>
- Pompilio, A., Bianconi, L., Regazzi, F., Bonora, P., 2005. RADAMES: A new management approach to opera: Repertory. archives and related documents, in: *Proceedings - First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution*. <https://doi.org/DOI:10.1109/AXMEDIS.2005.42>
- Popp, M.P., Dallis, D. (Eds.), 2012. *Planning and Implementing Resource Discovery Tools in Academic Libraries*, *Advances in Library and Information Science*. IGI Global. <https://doi.org/10.4018/978-1-4666-1821-3>
- Poria, S., Cambria, E., Ku, L.-W., Gui, C., Gelbukh, A., 2014. A Rule-Based Approach to Aspect Extraction from Product Reviews, in: *Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP)*. Presented at the Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, pp. 28–37. <https://doi.org/10.3115/v1/W14-5905>
- Race, T., 2012. Resource Discovery Tools: Supporting Serendipity, in: *Planning and Implementing Resource Discovery Tools in Academic Libraries*. IGI Global, pp. 139–151.
- Raimond, Y., Sandler, M., 2008. A web of musical information. Presented at the ISMIR 2008: Proceedings of the 9th International Conference of Music Information Retrieval, Lulu.com.
- Risé, C., 2006. *Don Giovanni, l'ingannatore*. Frassinelli, Milano.
- Roccatagliati, A., 1996. *Felice Romani librettista, Quaderni di Musica/realità*. Libreria musicale italiana, Lucca.
- Rojas, R., 1996. Unsupervised Learning and Clustering Algorithms, in: *Neural Networks*. Springer, Berlin-Heidelberg, pp. 99–121. https://doi.org/10.1007/978-3-642-61068-4_5
- Sartori, C., 1990. *I libretti italiani a stampa dalle origini al 1800*. Bertola e Locatelli, Cuneo.
- Seragnoli, D., 1987. La struttura del personaggio e della fabula, in: *Il Teatro Italiano Nel Rinascimento, Problemi e Prospettive. Serie Di Musica e Spettacolo*. Il Mulino, Bologna, pp. 297–317.
- Smiley, D., Pugh, E., Parisa, K., Mitchell, M., 2015. *Apache Solr enterprise search server*. Packt Publishing Ltd, Birmingham.
- Somerville, I., 2005. *Ingegneria del software*. Addison Wesley.
- Spink, A., Zimmer, M., 2008. *Web search: Multidisciplinary perspectives*. Springer Science & Business Media, Berlin-Heidelberg.
- Stara, A., 2004. *L'avventura del personaggio*. Le Monnier università, Firenze.
- Sudeepthi, G., Anuradha, G., Babu, M.S.P., 2012. A Survey on Semantic Web Search Engine. *IJCSI International Journal of Computer Science Issues* 9.
- Tomasi, F., 2018. Modelling in the Digital Humanities: Conceptual Data Models and Knowledge Organization in the Cultural Heritage Domain. *Historical Social Research/Historische Sozialforschung. Supplement* 170–179.
- Tomasi, F., 2008. *Metodologie informatiche e discipline umanistiche*. Carocci.
- Troy, C.E., Weiss, P., 2001. *Intermezzo*. Grove Music Online.
- Viégas, F.B., Wattenberg, M.D., 2004. Studing cooperation and conflict between authors with history flow visualizations, in: *Proocedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, pp. 575–582.
- Vitali S., 2004. *Passato digitale: le fonti dello storico nell'era del computer*. Bruno Mondadori.

- W3C OWL Working Group, 2004. OWL Web Ontology Language.
- Weigl, D.M., Page, K.R., Lewis, D., Crawford, T., Knopke, I., 2015. Unified access to media industry and academic datasets: a case study in Early Music, in: Extended Abstracts for the Late-Breaking Demo Session. Presented at the 16th International Society for Music Information Retrieval Conference, Malaga.
- Windhager, F., Mayr, E., Schreder, G., Smuc, M., Federico, P., Miksch, S., 2016. Reframing Cultural Heritage Collections in a Visualization Framework of Space-Time Cubes. *HistoInformatics@ DH* 20–24.
- Zamudio Lopez, S.A., Santaolaya Salgado, R., Fragoso Diaz, O.G., 2012. Restructuring Object-Oriented Frameworks to Model-View-Adapter Architecture. *IEEE Latin Am. Trans.* 10, 2010–2016. <https://doi.org/10.1109/TLA.2012.6272488>